Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
NEAL E. MILLER, *Yale University*
KENNETH W. SPENCE, *State University of Iowa*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Assistant Managing Editor
Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS, SARAH WOMACK

CONTENTS

A Test of a Statistical Learning Theory Model for Two-Choice Behavior With Double Stimulus Events: N. H. ANDERSON AND D. A. GRANT 305

Reward and Punishment in a Minimal Social Situation: J. B. SIDOWSKI 318

Effects of Varied Distance on Short-Range Size Judgments: N. JENKIN 327

The Interaction of Association Value and Stimulus Configuration in Size Estimation: M. E. DOW AND J. E. GORDON 332

Semantic Generalization With Experimentally Induced Associations: W. E. JEFFREY AND R. J. KAPLAN 336

The Immediate Effects of Emotionality Upon Behavior Strength: N. R. ELLIS 339

Emotionality and the Yerkes-Dodson Law: P. L. BROADHURST 345

Transfer From Verbal Pretraining to Motor Performance as a Function of Response Similarity and Angle of Movement: D. R. HOFFELD 353

The Discrimination of Speech Sounds Within and Across Phoneme Boundaries: A. M. LIBERMAN, K. S. HARRIS, H. S. HOFFMAN, AND B. C. GRIFFITH 358

Retention of Verbal Associates as a Function of Strength: W. N. RUNQUIST 369

Supplementary Report: Effect of Verbal Pretraining on the Acquisition of a Complex Motor Skill: W. F. BATTIG, D. R. HOFFELD, S. SEIDENSTEIN, AND W. J. BROGDEN 375

American Psychological Association

Vol. 54 No. 5 November 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the *Publication Manual of the American Psychological Association*. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 54, NO. 5 NOVEMBER, 1957

A TEST OF A STATISTICAL LEARNING THEORY MODEL FOR TWO-CHOICE BEHAVIOR WITH DOUBLE STIMULUS EVENTS [1,2]

NORMAN H. ANDERSON[3] AND DAVID A. GRANT

The University of Wisconsin

This paper reports three experiments bearing on a mathematical model for two-choice behavior based on statistical learning theory (**3, 4**). The model is closely related to that used by Estes and Straughan (**5**) for the case in which exactly one of the two choices is correct on each trial. They assume the existence of a population of stimulus elements whose defining properties are as follows. Each element is either in the state of being conditioned to the response class A_1 (predicting event E_1), or else in the state of being conditioned to response class A_2 (predicting event E_2). On each trial, the signal to respond activates a random sample containing the mean proportion, θ, of the population of stimulus elements. The response is then determined probabilistically according to the proportion of elements in the sample which are in each of the two states. At the end of the trial, all elements in the sample are conditioned to the response class which was in fact correct on that trial, independent of *S*'s actual response. The sample is then returned to the population and the process is repeated with a new random sample on the next trial, although there appears to be some necessity for assuming that successive samples are not independent.

The model tested here extends the above situation by allowing the occurrence, in addition to E_1 and E_2, of both E_1 and E_2, or neither E_1 nor E_2, on various trials. The occurrence of neither E_1 nor E_2 will be denoted as event E_0. The occurrence of both E_1 and E_2 will be denoted as event E_3. E_0 and E_3 will be called double events; E_1 and E_2 will be called single events. A number of experiments (**2, 9, 12, 13, 14**) have been performed with this experimental situation, and mathematical models for the case where E_0, E_1, and E_2, but not E_3, may occur, have been proposed by Neimark (**12**), and by Atkinson (**2**). Assuming that conditioning takes place in the way assumed in the model used by Estes and Straughan on trials where

[1] This paper is based on the first author's M.S. thesis (**1**) which was greatly facilitated by a National Science Foundation fellowship.

[2] Thanks are due to Nancy Angrist Myers who was a co-worker in the conducting of Exp. 2.

[3] Now at Yale University.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the *Publication Manual of the American Psychological Association*. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 54, NO. 5 NOVEMBER, 1957

A TEST OF A STATISTICAL LEARNING THEORY MODEL FOR TWO-CHOICE BEHAVIOR WITH DOUBLE STIMULUS EVENTS [1,2]

NORMAN H. ANDERSON[3] AND DAVID A. GRANT

The University of Wisconsin

This paper reports three experiments bearing on a mathematical model for two-choice behavior based on statistical learning theory (**3, 4**). The model is closely related to that used by Estes and Straughan (**5**) for the case in which exactly one of the two choices is correct on each trial. They assume the existence of a population of stimulus elements whose defining properties are as follows. Each element is either in the state of being conditioned to the response class A_1 (predicting event E_1), or else in the state of being conditioned to response class A_2 (predicting event E_2). On each trial, the signal to respond activates a random sample containing the mean proportion, θ, of the population of stimulus elements. The response is then determined probabilistically according to the proportion of elements in the sample which are in each of the two states. At the end of the trial, all elements in the sample are conditioned to the response class which was in fact correct on that trial, independent of *S*'s actual response. The sample is then returned to the population and the process is repeated with a new random sample on the next trial, although there appears to be some necessity for assuming that successive samples are not independent.

The model tested here extends the above situation by allowing the occurrence, in addition to E_1 and E_2, of both E_1 and E_2, or neither E_1 nor E_2, on various trials. The occurrence of neither E_1 nor E_2 will be denoted as event E_0. The occurrence of both E_1 and E_2 will be denoted as event E_3. E_0 and E_3 will be called double events; E_1 and E_2 will be called single events. A number of experiments (**2, 9, 12, 13, 14**) have been performed with this experimental situation, and mathematical models for the case where E_0, E_1, and E_2, but not E_3, may occur, have been proposed by Neimark (**12**), and by Atkinson (**2**). Assuming that conditioning takes place in the way assumed in the model used by Estes and Straughan on trials where

[1] This paper is based on the first author's M.S. thesis (**1**) which was greatly facilitated by a National Science Foundation fellowship.

[2] Thanks are due to Nancy Angrist Myers who was a co-worker in the conducting of Exp. 2.

[3] Now at Yale University.

single events occur, the critical point in the model is the disposition of the random sample on trials where double events occur. If E_2 occurs, it is hazardous to assume that the sample elements become conditioned to the response just made since, although it was correct, S can see that the other choice would have also been correct. A similar argument holds if E_0 occurs. Accordingly, a sharing parameter, α, is introduced for each double event which specifies what proportion of the sample is conditioned to the response just made (or, equivalently, on what proportion of the trials the entire sample is so conditioned).

Proceeding in the same way as in the Estes-Burke-Straughan (**4, 5**) development, one obtains the equation of the mean learning curve[4]

$$p(n) = p(\infty) - [p(\infty) - p(0)](1 - \beta)^n, \quad (1)$$

where

$p(n)$ = probability of response A_1 on trial n,

$p(0)$ = initial probability of response A_1

$$p(\infty) = \frac{\pi_0(1 - \alpha_0) + \pi_3(1 - \alpha_3) + \pi_1}{1 + \pi_0(1 - 2\alpha_0) + \pi_3(1 - 2\alpha_3)}$$

$$1 - \beta = (1 - \theta) + \theta[\pi_0(2\alpha_0 - 1) + \pi_3(2\alpha_3 - 1)]$$

and where

π_i = probability of event E_i, $i = 0, 1, 2, 3$

α_j = sharing parameter associated with event E_j, $j = 0, 3$.

It is seen that this equation is again a growth curve with asymptote, $p(\infty)$, depending on the π_i and α_j, and with learning rate, β, depending on the π_i and α_j as well as on θ, which is the learning rate in the case where only the two single events occur.

[4] Mathematical details, as well as the treatment of the general case of r responses and 2^r events may be found in Anderson (**1**).

To obtain an empirical estimate of the learning rate, knowing $p(\infty)$ and $p(0)$, Equation 1 is summed over N trials to yield

$$\sum_{n=0}^{N-1} p(n) = Np(\infty) - [p(\infty) - p(0)](1 - \beta)\frac{1 - (1 - \beta)^n}{1 - (1 - \beta)}. \quad (2)$$

Using the data to estimate the sum on the left, $1 - \beta$ is obtained by successive approximations.

If the α_j chance to assume the value unity, one obtains

$$p(n) = \frac{\pi_1}{\pi_1 + \pi_2} - \left[\frac{\pi_1}{\pi_1 + \pi_2} - p(0)\right] \times [1 - \theta(\pi_1 + \pi_2)]^n \quad (3)$$

which is independent of the individual values of π_0 and π_3, although not of their sum.[5] For $\pi_1 + \pi_2 = 1$, so that neither of the double events, E_0 or E_3, occurs, this reduces to the equation of Estes and Straughan.

Using the notation $p(x|y)$ to denote the probability of x given y, and considering only trials on which one of the single events occurs, the quantity a defined by

$$a = p(A_1|E_1) - p(A_1|E_2) \quad (4)$$

is an estimate of θ within the model, as was shown by Estes and Straughan. Without reference to the model, one may interpret a as a measure of stimulus control over S. Its values lie between -1 and $+1$. It assumes the value 0 if S's actions are independent of the stimulus sequence, the value $+1$ if S always makes next that response which was correct on the preceding trial, and the value -1 if S always makes next that response which was incorrect on the preceding trial.

Similarly, conditional probabilities of response may be considered for those

[5] It should perhaps be specifically noted that this is the appropriate generalization of the "matching solution" model (for these values of the sharing parameters). For example, if the event probabilities are .7 and .3, and if the events are *independent*, then the asymptote predicted by the model is .84, not .70.

trials on which the double events occur. Defining b as

$$b = p(A_1 \text{ on trial } n+1 \mid E_j \text{ and } A_1 \text{ on trial } n) - p(A_1 \text{ on trial } n+1 \mid E_j \text{ and } A_2 \text{ on trial } n), j = 0, 3, \quad (5)$$

then, within the model, b is an estimate of $\theta(2\alpha_j - 1)$, and if $\alpha_j = 1$, this gives another estimate of θ. Without reference to the model, b can be interpreted as a measure of perseverative tendency in the face of a noninformative event. Its values lie between -1 and $+1$. If S has alternation tendencies, then this quantity will be negative; if S has perseverative tendencies, then the difference will be positive.

Three experiments are reported here. The first is a pilot study which was run to gain preliminary information about behavior in this situation. The second is primarily methodological, investigating the effects of instructions and trial rate. The third experiment was designed specifically to test the model; its design follows closely the experiment of Estes and Straughan (5), and includes two "control" groups whose function is to calibrate the procedure used here against their results.

Method

Apparatus

The apparatus, slightly adapted from that used in previous work (11, 14), was the same for the three experiments. Two 6-w. opal lights, mounted 10 in. apart horizontally on a vertical, flat black, 15 × 20-in. panel, formed the stimulus display. These lights were activated by a punched tape fed through a Western Union tape transmitter. An audio oscillator generated an 800-cycle tone which served as *S*'s signal to respond. A control mechanism consisting of a 6-rpm synchronous motor and associated relays, cams, and microswitches governed the presentation of the tone and the sequence of stimulus lights.

A row of four partially separated *S*s' booths was placed 9 ft. from the stimulus display. Each booth contained a small box with two nonself-releasing keys, corresponding to the two lights, which *S*s pushed down and then up in making their predictions. Responses were recorded automatically with an Esterline-Angus recorder.

Except for the stimulus display and *S*s' booths, all equipment was placed on *E*'s desk which sat behind *S*s' booths and from which *E* monitored the course of the experiment.

Procedure

The *S*s were volunteers from elementary psychology classes who obtained class points for their services. They were randomly assigned, in groups of two, three, or four, to the various experimental conditions. No *S* was used more than once.

In each experiment, *S*s were seated in the booths and given instructions stating that *S* should always press *exactly one key* each time the tone sounded in order to indicate which light he thought would flash on next; that if both lights flashed, either choice would be counted correct; but that, if neither light flashed, this would not be counted against *S*.

Except for a few preliminary trials in Exp. 2 and 3, the successive events of the stimulus sequence were given without rest. The events occurred randomly, according to the probabilities designated below, subject to balancing in blocks of 20 trials for Exp. 1 and 2, and in blocks of 50 trials for Exp. 3.

Four independent event sequences were used in Exp. 1, and three independent sequences in each different probability condition of Exp. 2 and 3. In Exp. 1 and 2, the left light was always the more frequent. In Exp. 3, right and left were counterbalanced, and double events occurred on the same trials for corresponding groups.

Three experiments differ in instructional detail, length and composition of the stimulus sequences, and in trial rate as follows.

Experiment 1.—Following 12 initial trials in which each of the three possible stimulus events occurred four times in random order, 220 trials were given with these event probabilities: both lights, .5; left light alone, .4; and right light alone, .1. The trial duration was 10 sec. divided as follows: signal tone, 1 sec.; waiting time (during which responses were to be made), 3 sec.; stimulus lights on for 3 sec.; intertrial interval, 3 sec. Instructions, which were read by *E*, were similar to Instructions D of Exp. 2. There were 35 *S*s in this experiment.

Experiment 2.—The main variables were trial rate and instructions. The four sets of instructions, denoted by A, B, C, and D, had identical initial parts which gave operating instructions and told *S*s that their job was to make correct predictions. The four sets differed in their concluding parts which were, in brief: A: your

job is to outguess the experimenter; B: no additional content (besides the same exhortation to do well that was contained in the other three sets); C: this is an experiment in learning; you should try to learn how to improve the accuracy of your predictions; D: this is an experiment in learning; you should try to learn how to improve the accuracy of your predictions; there is no fixed pattern in the way the lights come on; act sensibly; if you can't get them all, get as many as you can. All instructions were presented with a tape recorder.

Each trial was divided thus: signal tone (during which responses were to be made), 2 sec.; waiting time, 1.2 sec.; stimulus event, .8 sec.; intertrial interval, 1, 3.5, or 6 sec. These values of the intertrial interval gave trial rates of 5, 7.5, and 10 sec. Groups run at these speeds received 360, 300, and 240 trials, respectively.

The experiment splits into two parts, denoted by Exp. 2.1 and Exp. 2.2, according to the event probabilities used. The two parts were run simultaneously. Experiment 2.1 was a 4×3 design formed from the four sets of instructions and the three trial rates. There were 234 *Ss*, 18 in each cell, except that 36 *Ss* were run using Instructions C and the 10-sec. trial rate. Eighteen of these were chosen at random for the analysis, but all 36 are used in the graphs. Event probabilities were: left light, .90; right light, .50.

Experiment 2.2 was a 1×3 design formed from the three trial rates and Instructions C. A total of 71 *Ss* was used. Event probabilities were: left light, .75; right light, .50.

All *Ss* in Exp. 2 received these four preliminary trials: left light alone, right light alone, neither light, both lights, in that order. Following these trials, the apparatus was stopped and *Ss* were asked if they had any questions. Questions pertaining to the nature or purpose of the experiment were answered by paraphrasing the instructions.

TABLE 1

GROUP DESIGNATIONS AND EVENT PROBABILITIES OF TRIALS 101–400 OF EXP. 3

Group	π_0	π_2	π_1	π_1
.15	0	0	.15	.85
.30	0	0	.30	.70
.30–0	.30	0	.15	.55
.30–3	0	.30	.15	.55
.45–0	.45	0	.15	.40
.45–3	0	.45	.15	.40
.60–0	.60	0	.15	.25
.60–3	0	.60	.15	.25

Note.—π_i = Probability of Event E_i.

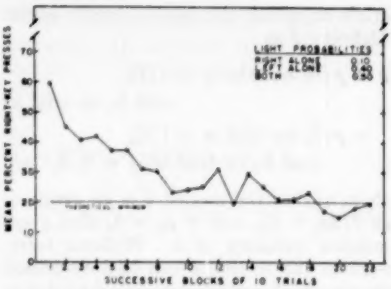

FIG. 1. Mean percentage of right-key presses as a function of trial blocks in Exp. 1.

Experiment 3.—All eight groups in Exp. 3 were run using the 5-sec. trial rate and Instructions C of Exp. 2 (except for small changes in the wording for the two groups receiving only single events). Instructions were read by *E*. Six groups formed a 3×2 design with probability (.30, .45, .60) and type (neither light, both lights) of double event as variables. Two groups received only the two single events (right light alone, left light alone) and served as control groups to calibrate the present procedure against that of Estes and Straughan (**5**).

All groups were given three preliminary trials: both lights, neither light, both lights, in that order. The apparatus was then stopped and questions were answered as in Exp. 2. There followed 424 trials without pause. The first 24 trials were adaptation trials in which the two or three events appropriate to each group appeared equally often. All groups were treated alike ($\pi_0 = 0$, $\pi_3 = 0$, $\pi_1 = .75$, and $\pi_2 = .25$) on the next 100 trials in which only single stimulus events occurred. The groups were treated differently on the last 300 trials, but in all cases, the light previously the more frequent became the less frequent. The event probabilities and group designations for the last 300 trials are given in Table 1. Right and left were counterbalanced so that E_1 refers to the event occurring .75 of the time on the first 100 trials.

RESULTS

Experiment 1.—The data of Exp. 1 are summarized in Fig. 1 which plots mean percentage of right-key presses as a function of trial number, omitting the 12 initial trials. The horizontal line at 20% is the minimum asymptote allowed by the model, corresponding to a value of the sharing parameter,

$\alpha_3 = 1$. The mean curve starts significantly (.05) above the 50% level and descends to the neighborhood of the theoretical minimum. Accordingly, α_3 would be expected to have a value close to 1. At the same time, however, the terminal performance suggests that the curve may still be decreasing, that asymptotic performance has not been reached, and, consequently, that the model does not describe the data. Therefore, it was decided to perform Exp. 2 primarily to test effects of instructions and trial rate in order to obtain more information as to the limits of applicability of the model.

Experiment 2.—Figure 2 plots mean percentage of right-key presses as a function of trial number for Exp. 2. The lower three curves represent the data of Exp. 2.1 pooled over instructions which had negligible effect. Starting somewhat above the 50% level, the curves descend steeply to an apparently stable level a little below 10%, with the 5-sec. group doing poorest, the 7.5-sec. group best, and the 10-sec. group lying between. Over Trials 181–240, the response rates are 9.8, 6.3, and 7.9% for the 5-, 7.5-, and 10-sec. groups, respectively. The three upper curves represent the data of Exp. 2.2. Here the curves fall in a natural order with the faster trial rate groups doing poorer. Mean response rates over Trials 181–240 are 32.6, 26.8, and 18.8% with standard errors (based on different N's) of 4.1, 3.7 and 1.5% for the 5.-7.5-, and 10-sec. groups, respectively,

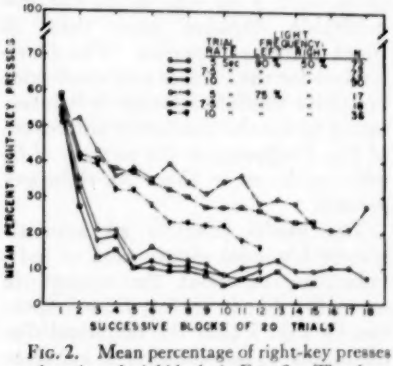

FIG. 2. Mean percentage of right-key presses as a function of trial blocks in Exp. 2. The three upper curves are from Exp. 2.2; the three lower curves are from Exp. 2.1.

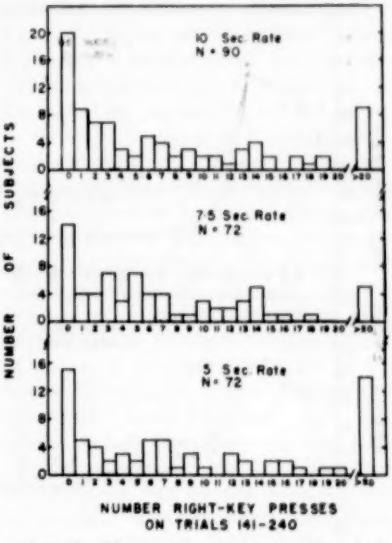

FIG. 3. Histograms of the number of *Ss* making given numbers of right-key presses on Trials 141–240, inclusive, for the three trial rates of Exp. 2.1.

Analysis of variance on Exp. 2.1 of total scores on Trials 1–240 showed that only trial rate was significant (.05). It had been intended to perform a similar analysis over Trials 141–240 in order to test rate effects on asymptotic performance. However, the extreme skewing of the scores (see Fig. 3) precluded this, so the test was made on total scores over Trials 41–240. Nothing approached significance in this test.

Analysis of variance for Exp. 2.2 over Trials 1–240 showed that trial rate was significant (.05). A similar

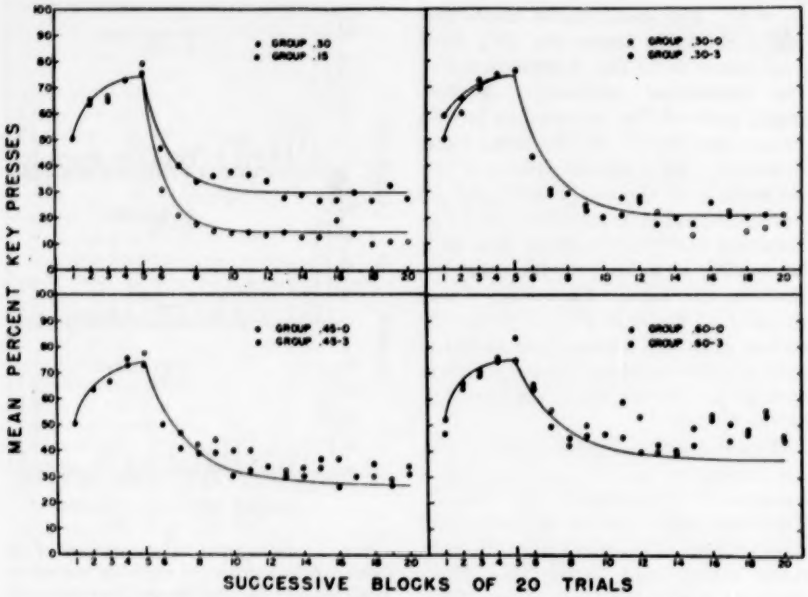

FIG. 4. Mean percentage of A_1 responses as a function of trial blocks in Exp. 3. The curves are theoretical; see text.

analysis over Trials 141–240 showed that trial rate was again significant (.01).

Although the experiment was not designed primarily to test agreement between terminal performance and theoretical asymptote, some information is available on this point. Performing *t* tests on mean scores of the three groups of Exp. 2.2 over the last 60 trials for each group showed that only the 10-sec. group departed significantly (.01) from the minimum theoretical asymptote obtained by setting both sharing parameters equal to unity.

Similar tests for Exp. 2.1 are difficult to make because of the skewed distributions of the scores. However, more incisive information can be gained by considering these very distributions which are plotted in Fig. 3. This figure lists, for each of the three rate conditions, the number of *Ss* who made given numbers of right-key presses in Trials 141–240, inclusive. For example, in the 10-sec. condition, 20 *Ss* did not press this less profitable key even once in these 100 trials while there were 9 *Ss* who made the less profitable response more than 20 times in the same period. The distributions for the various rate conditions are quite similar although it is interesting to see the consistent differences of Fig. 2 reflected in the number of *Ss* who made more than 20 right-key presses.

The model predicts an approximately binomial distribution of individual scores about the asymptotic mean (5), but it is clear from inspection of Fig. 3 that the empirical distributions do not conform to expectation. Incomplete χ^2's, based on the contribution of the single leftmost

rectangle, already gave a significant (.01) departure from binomiality for each of the three histograms.

Experiment 3.—The mean learning curves for the eight groups are shown in Fig. 4 as a function of trial number, ignoring the 24 adaptation trials. The groups start from about the 50% level on the first block of trials, increase to about the 75% level by Trial 100, reverse their preferences as the formerly less frequent light becomes the more frequent, and then settle down to a more or less steady rate of responding.

Two analyses of variance were performed on the 3×2 design of this experiment, one for total A_1 responses over Trials 101–400, one for total A_1 responses over Trials 301–400. Right vs. left, having been kept orthogonal to the two main variables, was included as a factor in the analysis. In both tests, only frequency of double event was significant (.01). Neither of the other variables, nor any of the interactions approached significance. In particular, it may be concluded that type of double event had little or no effect on number of A_1 responses. The error terms from these analyses are used in the tests of the theoretical curves given in Table 3.

The theoretical curves of Fig. 4 were fitted as follows. Estimates of θ for each group were obtained using Equation 2 transformed to yield an equation similar to Equation 6 of Estes and Straughan (**5**). Here $p(\infty) = \pi_1 = .75$, and $\beta = \theta$, since only single events are involved. The initial probability was estimated from the first block of 20 trials. Estimates of the sharing parameters were obtained by substituting the asymptotic response rates of .32 and .30 for Groups .45–0 and .45–3, respectively, together with the appropriate values of the π_i in the expression for $p(\infty)$ below Equation 1. This yielded the estimates, $\alpha_0 = .84$, and $\alpha_3 = .92$. These values are close to the unit values suggested by the previous experiments, and by the work of Neimark (**12**). For simplicity, each sharing parameter was set equal to 1, a procedure which does not materially affect the over-all goodness of fit considerations.

With values of θ and α_j at hand, Equation 1 was used to plot the theoretical curves over Trials 101–400 by starting counting on n at Trial 101. The initial probability, $p(0)$, over this stretch was given the a priori value of .75, the theoretical asymptote over the first 100 trials. Since the curves for the paired groups never differed by

TABLE 2

CONSTANTS OF EQUATION 2 USED TO CALCULATE THEORETICAL CURVES FOR EXP. 3

Group	Trials 1–100			Trials 101–400			Learning Rates Estimated from Trials 101–200	
	$p(0)$	$p(\infty)$	θ	$p(0)$	$p(\infty)$	$\beta = [\theta(+)]$	$\beta = [\theta(+)]$	θ
.15	.51	.75	.045	.75	.150	.045	.107	.107
.30	.50	.75	.036	.75	.300	.036	.048	.048
.30–0	.50	.75	.039	.75	.214	.027	.067	.095
.30–3	.59	.75	.036	.75	.214	.025	.068	.097
.45–0	.51	.75	.040	.75	.273	.022	.037	.067
.45–3	.51	.75	.037	.75	.273	.020	.037	.067
.60–0	.52	.75	.053	.75	.375	.021	.026	.066
.60–3	.47	.75	.064	.75	.375	.026	.028	.070

TABLE 3

COMPARISON OF PREDICTED AND OBSERVED MEAN TOTAL A_1 RESPONSES FOR EXP. 3

Group	Trials 301–400			Trials 101–400			Predicted from New θ Values
	Obs.	σ_m	Pred.	Obs.	σ_m	Pred.	
.15	13.1	1.46	15.0	46.3	4.54	58.3*	50.6
.30	29.1	1.32	30.0	99.1	5.46	102.6	99.3
.30–0	20.0	2.56	21.4	70.7	6.40	83.9*	72.2
.30–3	21.4	2.56	21.4	71.1	6.40	85.6*	72.0
.45–0	31.9	2.56	27.3	111.0	6.40	103.7	94.9*
.45–3	29.9	2.56	27.3	102.1	6.40	105.5	94.8
.60–0	49.4	2.56	37.5**	146.5	6.40	130.4*	125.8**
.60–3	49.8	2.56	37.5**	145.0	6.40	127.2**	125.8**

* Predicted and observed significantly different, $P < .05$.
** Predicted and observed significantly different, $P < .01$.

more than 3% and were generally much closer, only the curves for the groups receiving E_0 are shown, except for Group .30–3 over Trials 1–100. The constants of the curves and the learning rates are summarized in Table 2.

A first test of agreement between predicted and obtained asymptotic levels of response was made by comparing predicted and obtained numbers of A_1 reponses over Trials 301–400. These tests are summarized in Columns 2–4 of Table 3. Only for Groups .60–0 and .60–3 is there a significant discrepancy. A more stringent test is the comparison of total responses over Trials 101–400 as predicted and observed. The results of these tests are given in Columns 5–7 of Table 3. The values are in the proper order for the most part but show large absolute discrepancies.

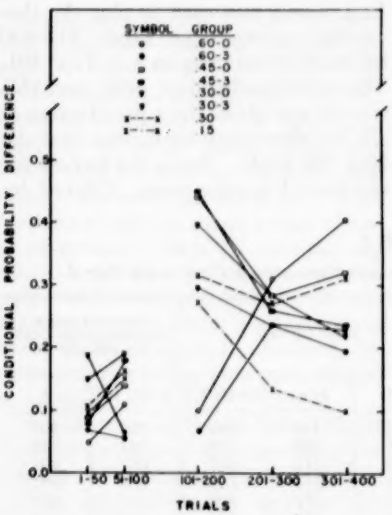

FIG. 5. Mean conditional probability difference of Equation 4 as a function of trial blocks in Exp. 3.

Inspection showed that the data points tended to fall below the curves on Trials 101–140 indicating that the θ values used were too small. To check on this, new values of θ, given in the last column of Table 2, were estimated from Trials 101–200 and used to calculate total responses on Trials 101–400. These values are given in the last column of Table 3. These second estimates of θ are uniformly larger than the original estimates, and the agreement between predicted and observed numbers of responses is considerably better.

As stated in the introduction, the difference of the conditional probabilities, a, of Equation 4 gives a measure of stimulus control over S's responding. Figure 5 plots a as a function of trial number. There is a general

modest increase from the first half to the second half of Trials 1–100 marred only by the anomalous behavior of Group .30–3. The groups diverge radically on Trials 101–200 with all except Groups .60–0 and .60–3 showing considerable increases. Group .30 then continues at a high level whereas Group .15 shows a considerable decrease in the later trial blocks. For the 3×2 design, the tendency is for the groups differing only on type of double event to be equal initially but to diverge pairwise so that on the last block of 100 trials, the E_0 groups tend to lie above the corresponding E_2 groups, the difference being near zero for Groups .30–0 and .30–3 and increasing with increasing frequency of double event.

These statements are made more precise by the application of Grant's (6) extension of Alexander's test for trend over the last three blocks of 100 trials, as summarized in Table 4. For the two control groups, Groups .15 and .30, line A indicates significant linear and quadratic components for the data of the two groups considered together. Line B shows that the means for these two groups differ over these 300 trials while Line C indicates that the groups differ in the slope of their curves.

The right-hand colums of Table 4 give the same analysis for the 3×2 design. Here a further breakdown into variance associated with the factors in the design is possible. Line A indicates that there is no marked trend in the data considered as a whole. Line B shows that the group means over these 300 trials do not differ significantly. Line C indicates significant differences in the linear and quadratic components of the group curves. For both components, group differences arise from frequency, but not type, of double event. At the same time, there is a significant source

TABLE 4

TREND ANALYSIS FOR CONDITIONAL PROBABILITY DIFFERENCE OF EQUATION 4 (SEE FIG. 5)

Source	Error Term	Groups .15 and .30			3×2 Design		
		df	*SS*	*F*	*df*	*SS*	*F*
A. Over-all trend	E	(2)	(.2479)	(6.91)**	(2)	(.0486)	.62
1. Linear	E_1	1	.1747	8.26**	1	.0393	.97
2. Quadratic	E_2	1	.0732	4.98*	1	.0093	.24
B. Between-group means	D	1	.6303	11.20**	(5)	(1.4449)	(1.67)
a. Frequency	D				2	.9141	2.64
b. Type	D				1	.4244	2.45
c. Interaction	D				2	.1063	.31
C. Between-group trends	E	(2)	(.1947)	(5.43)**	(10)	(2.8739)	(7.29)**
1. Linear	E_1	1	.1945	9.19**	(5)	(2.3805)	(11.73)**
a. Frequency	E_1				2	2.0719	25.52**
b. Type	E_1				1	.0514	1.27
c. Interaction	E_1				2	.2572	3.17*
2. Quadratic	E_2	1	.0002	—	(5)	(.4934)	(2.58)*
a. Frequency	E_2				2	.4745	6.20**
b. Type	E_2				1	.0090	.24
c. Interaction	E_2				2	.0099	.13
D. Between ind. means	E	46	2.5877		138	23.9318	4.40**
E. Between ind. trends		(92)	(1.6495)		(276)	(10.8854)	
1. Linear		46	.9732		138	5.6022	
2. Quadratic		46	.6764		138	5.2832	

* $P < .05$.
** $P < .01$.

TABLE 5

PERSEVERATIVE TENDENCIES FOR THE EXPERIMENTAL GROUPS OF EXP. 3, AS CALCULATED FROM EQUATION 5

Group	Trials					
	101–150	151–200	201–250	251–300	301–350	351–400
.30–0	.48	.47	.54	.62	.49	.44
.30–3	.50	.50	.45	.50	.38	.40
.45–0	.28	.39	.36	.47	.44	.48
.45–3	.32	.46	.42	.47	.61	.35
.60–0	.51	.56	.47	.42	.47	.46
.60–3	.40	.54	.51	.45	.44	.47

of variation in group slopes arising from the interaction of frequency and type of double event. This interaction, because of the lay of the points on Trials 101–200, reflects itself as an effect of type of double event in the last trial block which was found to be significant (.05) in an analysis of the scores for that block. Finally, Line D indicates significant individual differences in mean scores.

The second conditional probability difference, b, of Equation 5, gives a measure of perseverative tendency when responses are followed by the uninformative double events. The group data for b have been computed for the last six blocks of 50 trials, and are listed in Table 5. The b values are uniformly positive, of the order of .45, indicating that a strong perseverative tendency exists. There do not appear to be any marked differences between groups or between trial blocks.

Values of a and b were also computed for the 24 adaptation trials. The values of a ranged from $-.12$ to .07 with a mean of $-.01$ for the eight groups, a result which is consistent with the general increases in Fig. 5. The b value for the three E_0 groups combined was .12, which was significantly (.01) above 0 by a χ^2 test. For the three E_3 groups combined, the b value was .06 which was not significantly above 0. These values being positive, the implication is that there is little or no initial tendency to alternate from the last response.

DISCUSSION

Effect of trial rate.—In Exp. 2, a weak effect of trial rate was found with the more extreme light frequencies of 90% vs. 50%, while a pronounced rate effect was obtained for the less extreme frequencies of 75% vs. 50%. This is consistent with the results of earlier work (**7, 10**), in which no rate effects were found for 100% vs. 0%, and suggests that rate effects begin to appear as the situation is made more complicated.

Effect of instructions.—Failure to find any noticeable effect of instructions implies that the behavior is so much the more stable. Caution should be used in generalizing this conclusion since the results on trial rate raise the possibility that instructions may prove significant when the light frequencies are less extreme than the 90% vs. 50% used in Exp. 2.1.

Individual differences.—The wide departure of the distributions of Exp. 2.1 (Fig. 3) from expectation implies that the model in its present simple form does not apply to the case where relatively extreme values of the event frequencies are used. Furthermore, the results for the 5-sec. group which ended up very close to the theoretical asymptote of 10% underscore the remark of Estes and Burke (**4**) that agreement between theoretical and predicted learning curves is only a first and sometimes insensitive test of the adequacy of the model.

Loosely speaking, it would seem that those *S*s who gave no nonoptimal responses were acting rather differently from say, those who gave more than 20% nonoptimal responses. If so, it follows that a revised model should allow room for differences in individual strategies.

Number of trials.—Inspection of Fig. 4 indicates that Groups .60–0 and .60–3 tend to follow the theoretical curves ex-

cept for the last 100 trials. However, no comparable tendency to "break" appears in the data for the other six groups despite the large number of trials.

Asymptote predictions.—Except for Groups .60–0 and .60–3, satisfactory predictions of asymptotic performance in Exp. 3 were obtained by setting the sharing parameters equal to 1. For these two exceptional groups, the difference in the total frequencies of the two lights is only 10% and the results suggest a systematic error in the direction of too low a predicted asymptote when the differential light frequency is small.

Estimates of learning rates.—In Exp. 3, it had been hoped that the 24 initial trials would wash out any initial key preferences, negative recency effects, or alternation tendencies which may have existed so that $p(0)$ could be assigned the a priori value of .5, and so that the estimates of θ from Trials 1–100 would be applicable to Trials 101–400. Obviously this was not the case, and the readiness with which the groups reverse their key preferences after Trial 100 is in marked contrast to the reluctance with which they leave the 50% level at the beginning of the experiment. This increase in θ rendered the predicted curves for Trials 101–400 incorrect. When learning rates were estimated from Trials 101–200, considerably more satisfactory fits were obtained. While this indicates that the model will give a fairly good description of the mean learning curves as long as the differential light frequency is not too small, this is a weaker form of validation than had been hoped for.

Unfortunately, it is not clear from the present data how to get appropriate estimates of θ without using the very data to be predicted. The observed increases in θ may have arisen from continued adaptation to the situation, or from previous experience with a change in the identity of the more frequent light. A further possibility, indicated by Neimark (12), is that the value of θ increases with increases in the magnitude of the change in light frequencies. The present data are not inconsistent with this idea.

Conditional probabilities.—Perhaps the most interesting results are the differences of the conditional probabilities of response as given in Fig. 5 and Table 5.

Both the increase in a over the early blocks of trials, as well as the tendency for a to overestimate θ, which were found here, have been observed by Estes and Straughan. They conjectured that these effects arose from nonindependence of successive samples. Although this appears to be a reasonable interpretation for the data of Trials 1–100, it is not clear how it would handle the differences between group means and trends over Trials 101–400. Finally, it should be noted that the latter data indicate an effect of type of double event, although no sign of this appears in the mean learning curves.

According to the model, the b values should also give an estimate of θ. It is seen that they are also considerably too large. Moreover, they do not agree with the a values, either in average magnitude, or in change with trials.

Without reference to the model, these two conditional probability differences have straightforward psychological interpretations—a as a measure of stimulus control exerted by immediately preceding single events (possibly confounded, however, with the influence of more remote events), and b as a measure of perseverative tendency after double events. The results of Hake and Hyman (8) indicate that further information will be obtained when response probabilities conditional upon more remote stimulus events are considered. The present results suggest that such analyses give more precise knowledge of the immediate determinants of behavior, as well as more incisive tests of model and theory, than does the study of mean performance curves.

Finally, it should be noted that the b values may well depend on the physical components of the response system. Although little trouble was experienced with *S*s leaving one key down for several trials, it seems plausible that a different type of key system would give different results for this measure.

Other proposed models.—Groups .15 and .30 replicate the experiment of Estes

and Straughan (5), and similar results were obtained for the mean learning curves. The between-groups differences in mean and trend of *a* values which were found here arise largely from the last 200 trials, and the values for the earlier trials agree with the data from their shorter sequences. Hence, the results for these two groups confirm their experiment except for the later increases in learning rate and, possibly, the later behavior of the *a* values.

With respect to the model proposed here for the six experimental groups, the finding that the best estimate of α_0 is close to 1 is consistent with the model proposed by Neimark who made the a priori assumption that α_0 is equal to 1. Atkinson (2) has compared three models including the cases where $\alpha_0 = \frac{1}{2}$, and $\alpha_0 = 1$, as well as the case in which the stimulus sample is not conditioned to either response class when an E_0 occurs. This latter assumption, which is favored by his results, leads to exactly the same expressions for mean performance curves and *a* values for the event frequencies used here as does the assumption $\alpha_0 = 1$. Thus, although the mean curves do not differentiate, it should be possible further to test Atkinson's model against the *a* and *b* values which have been obtained here.

Statistical learning theory.—Although we conclude that the model is inadequate to meet the demands which we have imposed, this is not meant as an indictment of the underlying theory. The model has given a fair rationalization of the mean performance curves, and has failed principally in accounting for response distributions and conditional probabilities which are not even obtainable from other existing theories. The basic notions of statistical learning theory are concerned with a rigorization of the notions of stimulus and response, permitting the possibility of precise prediction. The partial success of the present model indicates more that the behavior under study is not as simple as the model assumed than that the basic theory is incorrect.

Summary

This paper reports a model for two-choice predicting behavior, and three experiments which evaluate the model. The model is based on statistical learning theory, being an extension of the Estes-Burke-Straughan model to the case where either response alone (single events), both responses, or neither response (double events) may be correct in any trial. Sharing parameters are introduced to represent the proportion of stimulus elements conditioned to the two response classes on each type of trial.

In each experiment, *S*s were required to predict which one of two lights would flash next by choosing on each trial *exactly one* of the two possible responses.

The results of Exp. 1 were consistent with the model but raised questions as to its limits of applicability.

Experiment 2 tested effect of four sets of instructions and three trial rates. Instructions had no discernible effect. Trial rate had a weak initial effect when the light frequencies were 90% vs. 50%, and a pronounced effect when the frequencies were 75% vs. 50%. Asymptotic response rates tended to lie at or below the theoretical minimum. Distributions of responses about the asymptotic mean departed widely from those predicted by the model.

Experiment 3 was designed primarily to test the model. Two groups received only single events as in the experiment of Estes and Straughan. The theoretical learning curves for these two groups generally substantiated their results. However, trend analysis of conditional probabilities of response indicated a possible discrepancy from their model.

Six additional groups formed a 3×2 design with type (neither response correct, both responses correct) and probability (.30, .45, .60) of double event as factors. Analysis of variance showed that only probability of double event had a significant effect on the mean learning curves. The model gave fair, but not perfect, predictions of terminal response rates and, when an increase in learning rate later in the trial sequence was taken into account, of the entire learning curves over Trials 101–400.

Two differences of conditional probabilities of response, one a measure of stimulus control over *S*, the other a measure of perseverative tendency after a double event, were also computed. Trend analysis of the former showed significant between-groups trends. Inspection of the latter showed little effect of group treatments or trials. Although, within the model, both differences give estimates of the learning rate, both were considerably larger than learning rates obtained from the learning curves. Nor did they agree

well with each other, either in magnitude or in change with trials.

It is concluded that the model should be useful in predicting mean performance over a fairly wide range of stimulus conditions but that it is inadequate to handle response distributions or conditional probabilities in its present form.

References

1. Anderson, N. H. Studies in two-choice predicting behavior. Unpublished M.S. thesis, Univer. of Wisconsin, 1955.
2. Atkinson, R. C. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, **52,** 28–32.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57,** 94–107.
4. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60,** 276–286.
5. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47,** 225–234.
6. Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.
7. Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the intertrial interval on the Humphreys' 'random reinforcement' effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40,** 609–612.
8. Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45,** 64–74.
9. Hornseth, J. P. Discrimination of frequency differences as a function of their relative and absolute values. Unpublished Ph.D. thesis, Univer. of Wisconsin, 1952.
10. Humphreys, L. G., Miller, J., & Ellson, D. G. The effect of intertrial interval on the acquisition, extinction, and recovery of verbal expectations. *J. exp. Psychol.*, 1940, **27,** 195–202.
11. Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49,** 343–351.
12. Neimark, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52,** 209–220.
13. Nystrom, C. O. Multiple-choice behavior as affected by knowledge of results and relative stimulus information. Unpublished Ph.D. thesis, Univer. of Wisconsin, 1954.
14. Peterson, R. M. The effect of probability and expected value on successive two-choice predictions. Unpublished Ph.D. thesis, Univer. of Wisconsin, 1954.

(Received September 26, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

REWARD AND PUNISHMENT IN A MINIMAL SOCIAL SITUATION [1]

JOSEPH B. SIDOWSKI

San Diego State College

The study of conditioning in the laboratory, and the development of conditioning theory, are justified in part by the hope that the principles derived will ultimately enhance the understanding of human behavior in complex situations including social ones. In the present study an attempt is made to re-evaluate the essential features of a social situation viewed entirely within the framework of conditioning theory, and to investigate the simplest situation that could be considered truly social within this framework. Analysis of social behavior in these terms implies that no assumption whatever is to be made regarding such concepts as attitude, understanding, etc. If such factors enter the picture at all, they are to be considered as manifestations of complex habits. Preferably, they will not be enountered at all in the minimal social situation, and in any event will not be part of the definition of social behavior.

In this analysis it will be assumed that the main factors controlling social behavior are reward and punishment. Within this framework, the essential features of a social situation are: (*a*) Two or more *S*s have at their disposal responses which result in rewarding or punishing effects on other *S*s. (*b*) The principal sources of reward and punishment for any *S* depend on responses made by other *S*s. (*c*) The responses controlling reward and punishment are subject to learning. It will be assumed that, whatever else may be involved, at least the above features are present in any social situation.

It is of interest to determine what learning would be predicted on the basis of these conditions alone, assuming only that the principles of conditioning theory are operating. This was the purpose of the present experiment.

Two *S*s isolated from each other were provided with push buttons by means of which each could give the other reward or punishment. In this situation, if *S* makes a response, he does not "suffer the consequences" directly. An *S* can change the rewards and punishments he receives only to the extent that this action may change the behavior of the other *S*. Thus, if *A* makes a response which results in a reward for *B*, *B* may be expected to continue with, or to repeat, his immediately preceding behavior; if *A* makes a response which results in a punishment for *B*, *B* may be expected to discontinue his immediately preceding activity. Such changes in *B*'s behavior would have a direct effect on the reward and punishment being received by *A*. In this way, *A* would receive some consequences from his own responses, but in an indirect way.

It is not intended that this situation be directly comparable to familiar everyday encounters. People generally enter social situations already equipped with a massive amount of prior learning relating to the effects of their behavior on others. This makes rigorous control of a given social experiment next to impossible.

[1] Based upon a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Wisconsin. The author wishes to express his gratitude to his adviser, Dr. D. A. Grant, and to Drs. L. B. Wyckoff and R. Miles for assistance in the preparation of this paper. The author also wishes to acknowledge the aid of the Numerical Analysis Laboratory of the University of Wisconsin.

However, much of social control of other people's behavior is through stylized ways of indicating approval and disapproval that are in effect methods of rewarding and punishing others. With adults, most of the rewarding and punishing techniques are conventional, but in many situations the responses learned are no longer appropriate. New responses must be learned. The process through which this learning and relearning takes place will be central to any extension of conditioning theory into the realm of social behavior. One may ask, Is face-to-face contact and "understanding" of the relationship between *Ss* necessary for "social" learning? The question remains as to whether learning will occur simply as a result of the operation of reward and punishment. Also, does it facilitate or hinder learning for the *S* to know that there is another person in the situation?

The first step in a program designed to answer these questions is an investigation of the course of learning in the minimal social situation. The *Ss* were divided into two major groups, an Informed and an Uninformed group. All *Ss* in the Informed group were told that another *S* was serving in the experiment. In the Uninformed group, the effects of previous social learning were minimized since *Ss* were not told (nor did they guess) that they were in a social situation at all. The two major groups were subdivided into three Reward-Punishment subgroups in which: (*a*) *Ss* could give each other Shock Only, (*b*) *Ss* could give each other Score Only, and (*c*) *Ss* could give each other both Shocks and Scores.

Method

Subjects.—The *Ss* were 180 members of elementary psychology courses at the University of Wisconsin who received class points for participating.

Apparatus.—The two experimental rooms were each equipped with a control panel which provided two push buttons, an electrically operated counter, and a pair of electrodes. Thus, pressing the right-hand button in Room 1 registered one point on the counter in Room 2, while pressing the left-hand button in Room 1 delivered an electric shock to the electrodes in Room 2. In Room 2 the right and left buttons controlled the shock and counter, respectively, in Room 1. A white light served as a signal that the experiment had begun. It remained on until the session ended. The *E*'s control board contained counters which indicated the number of times that each of the four buttons had been pressed. In addition to the counters, an Esterline Angus recorder was used to give continuous time record of each response, and of the four possible combinations of pairs of consecutive responses. For example, one pen of the recorder operated when a left button response by *A* was followed by a left button response by *B*; another pen recorded when a left button response by *A* was followed by a right button response by *B*, etc.

The shock was delivered through the plate circuit of a 6SK7 tube so that very high resistance was in series with *S*. This minimized effects of changes in *S*'s resistance. Separate controls were available for each of the two *Ss*. Shock duration was regulated by a relay controlled timing circuit which permitted a constant shock interval for each *S*.

Design and procedure.—A 2×3 factorial design was used in the experiment. The 180 *Ss*, 90 pairs, were randomly assigned to six groups consisting of 15 pairs each, with the limitation that each group consisted of 5 pairs of males, 5 pairs of females, and 5 mixed pairs. All of the Uninformed conditions were run first and all Informed conditions were run last to minimize accidental information to *S*. Within each of these conditions *Ss* were assigned at random to the subgroups.

In the Uninformed condition, *Ss* were not told that another *S* was serving in the experiment. In the Informed condition, *Ss* were told that another *S* was serving in the experiment and that their own scores and shocks were controlled by the other *S*. Within the Uninformed and Informed conditions there were three subconditions. Under the Shock Only condition, the score circuit was inoperative but the shock circuit remained in operation. Subjects under the Shock Only condition, therefore, received no scores. Under the Score Only condition, the shock circuit was inoperative so that *Ss* could receive scores only. Under the Shock and Score condition, *Ss* were able to receive and administer both shocks and scores.

The amount of shock which *S* received was a function of his own threshold. A threshold circuit was included in the apparatus in order to test the threshold of each *S* before entering the experimental rooms. Preliminary tests indicated that the absolute threshold was more reliable and clearly defined than a "pain" threshold, so this measure was used as a basis of shock levels. Three readings of each *S*'s absolute threshold were taken and the mean of his three readings was used as the threshold for each *S*. An analysis of variance was carried out on the thresholds of the six groups and no significant difference was found. A previous study (**4**) showed that strong shock was more conducive to learning in the minimal social situation than weak shock. Therefore, three times the absolute threshold in milliamperes was used as the shock level in this experiment. Any questions concerning the current were ignored; however, each *S* was offered the opportunity to withdraw from the experiment before the session began. One *S* withdrew.

After the threshold readings were completed, *S* was taken to one of the two experimental rooms where he was seated before the *S* control board. Electrodes were attached to the first two fingers of his left hand. Instructions were then read to *S* after which he was told to relax for a few minutes while the apparatus warmed up. During this time, *E* brought the second *S* before the *E* panel and went through the same procedure. The second *S* was taken to Room 2 and seated before the second control board. Instructions were then read to the second *S*.

All *S*s were given the following instructions:

"When the white light goes on the experiment begins. The experiment is in progress as long as the white light stays on. You can press either of these buttons in any manner that you wish, and as frequently as you wish. Do not attempt to press both buttons at the same time and use only your right hand for pressing. You are to keep your left hand on the table. Do not hold a button down, but just press and release it. The object of the experiment is to make as many points as you can. Your point score will appear on this counter, and the counter will turn each time that you score a point. When the white light goes off, it is a signal that the experiment has ended. Remain seated until *E* comes into the room. Remember, you are to try to make as high a score as you can in any manner that you can."

In addition to the above instructions, the Informed groups were told the following:

"There is another *S* in another room who controls the number of shocks and the number of scores which you will receive. You in turn control the number of shocks and scores which the other *S* receives."

Under no condition was any mention made by *E* of the function of the buttons or of the electrodes. Questions concerning anything other than the clarity of the instructions were ignored. When *S* asked about the instructions, the instructions were merely repeated.

At the conclusion of the experimental session, a series of questions was asked in order to determine whether or not any *S*s in the Uninformed groups were aware of the presence of another *S* in the experiment. Subjects in all groups were questioned concerning the function of the buttons, electrodes, and experiment. Responses to the questions at the end of the experiment indicated that none of the *S*s in Uninformed groups was aware that a second *S* was involved in the experiment.

The experimental session lasted for 25 min.

Results

In the present study, the dependent variable was defined as the number of button-pressing responses (Shock and Score) made by each *S*. The performance of each *S* was recorded in terms of the number of Shock and Score responses per minute.

In order to clarify the analysis, the results of the experiment are presented in two sections. (*a*) In the initial analysis, Score and Shock responses are reported over the entire 25-min. session in blocks of 5 min. each. (*b*) A finer analysis is presented on the

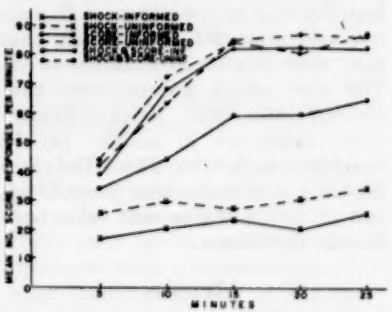

FIG. 1. Mean number of Score button responses per minute over blocks of 5 min. each for the 25-min session.

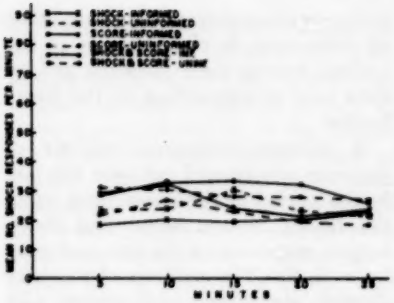

FIG. 2. Mean number of Shock button responses per minute over blocks of 5 min. each for the 25-min. session.

single minutes of the first 5-min. period and data are presented on some of the conditional dependencies in the sequences of responses in order to aid in the interpretation of the results.

Results of 25-min. analysis.—The data over the 25-min. session are plotted in Fig. 1 and 2. Figure 1 shows the mean number of Score responses per minute for each of the six experimental groups while Fig. 2 shows the mean number of Shock responses per minute. These two types of response were analyzed by means of Grant's extension of the Alexander trend test (**2**).

Inspection of the curves of Fig. 1 indicates that the mean number of Score responses increases over the 25-min. period for all except the two Shock Only groups. Further, each Uninformed group seems to do better than the corresponding Informed group.

More precise information on the

TABLE 1

SUMMARY ANALYSIS OF TRENDS FOR SCORES OVER 25-MIN. PERIOD

Source of Variation	*df*	*MS*	Error Term	*F*
A. Over-all trend	(4)	679,683	E	37.89**
1. Linear	1	2,179,315	E 1	56.37**
2. Quadratic	1	491,794	E 2	32.96**
3. Cubic	1	29,371	E 3	2.91
4. Quartic	1	18,252	E 4	2.26
B. Between-group means	(5)	2,087,920	D	8.47**
a. Information	1	586,143	D	2.38
b. Reward-Punishment	2	4,737,484	D	19.23**
c. Inf. × R-P	2	189,244	D	.77
C. Between-group trends	(20)	50,958	E	2.84**
1. Linear	(5)	145,627	E 1	3.77**
a. Information	1	17,710	E 1	.46
b. Reward-Punishment	2	354,162	E 1	9.16**
c. Inf. × R-P	2	1,051	E 1	.03
2. Quadratic	(5)	45,346	E 2	3.04*
a. Information	1	3,269	E 2	.22
b. Reward-Punishment	2	94,450	E 2	6.33**
c. Inf. × R-P	2	17,281	E 2	1.16
D. Between-ind. means	174	246,388	E	13.73**
E. Between-ind. trends	(696)	17,940		
1. Linear	174	38,661		
2. Quadratic	174	14,921		
3. Cubic	174	10,107		
4. Quartic	174	8,072		
F. Total	899			

* $P = .05$.
** $P = .01$.

rate of Score responding is given by the trend shown in Table 1. The existence of learning is shown by the significance of the linear and quadratic components of the over-all trend. The significance of the linear and quadratic components of between-groups trend shows that the learning is not the same for all groups. There were no group differences between cubic and quartic components which have been omitted from Table 1. A further breakdown into the factors of the orthogonal design shows that this difference in trends is associated with the three Reward and Punishment conditions. A similar statement holds for the analysis of the group means: significant differences arise only from the Reward and Punishment conditions. The apparent superiority of the Uninformed groups is not substantiated. Finally, the significant differences between individual means show merely that there are reliable individual differences.

Since significant differences between group means were found, Duncan's test (1) for differences between ranked means was carried out on the Score data. The differences between the two Shock Only conditions and all other conditions were significant at the .01 level. No significant difference was found between the Shock Only-Informed and the Shock Only-Uninformed groups, nor were the remaining four score conditions significantly different. The reward (Score), therefore, seems superior to punishment (Shock).

A trend analysis (2) was also used to test the Shock response data. Only the quadratic component of the over-all trend was significant. This suggests the presence of curvilinearity or "bowing" in the rate of responding to the Shock button as a function of time. A lack of significance between group means and trends indicated that all differences in the behavior of the various groups were characterized by their rate of responding to the Score button.

A repeated-measures analysis of variance was carried out over the last 5-min. period of the experiment using the Shock button score and Score button responses as the two measures for each *S*. The analysis of variance showed significant differences (.01 level) between groups, which arose from the influence of the Reward-Punishment condition. A significant Button effect reflected the greater number of responses made to the Score button. Finally, a significant Groups by Button interaction showed that the difference in rate of responding to the two buttons was different for the various groups. The application of Duncan's test (1) showed that the two Shock Only groups gave significantly fewer score responses than the remaining four groups, but no further differentiation of the groups was found. The same analysis, when applied to the Shock responses, showed no significant variation between the six groups.

Results of the analysis of first 5 min. and of sequences of responses.—Analysis of the data for single minutes of the first 5-min. period showed that learning did take place during this time. Trend analyses of the score responses showed significant between-group trends (linear component) and significant between-group means with the differences being due to differences between Reward-Punishment conditions. Analysis of the Shock button responses showed a significant quartic component suggesting the presence of reliable temporal fluctuations or oscillations in rate of responding over this initial period.

On a priori grounds it was thought

that learning would occur according to the following principles: (*a*) If S_A presses the Score button, S_B will tend to keep pressing whichever button he pressed previously; (*b*) If S_A presses the Shock button, S_B will tend to change buttons. It is easily seen that, singly or in combination, Principles *a* and *b* will tend to lead to Score responses for both *S*s. Hence, there should be an increase in Score-Score responses, a decrease in Shock-Score and Shock-Shock, and possibly an initial increase in Score-Shock followed by a decrease. To see if this happened, sequences of responses by pairs of *S*s were recorded and plotted for each of the six groups over the entire 25-min. session and also over the first 5-min. interval.

Figure 3 shows the mean number of sequential responses per minute for the Shock Only-Informed group over the 25-min. session. It is representative of the types of sequential curves found for the Shock Only conditions over the entire session and over the first 5 min. Figure 4 shows the sequential responses for the Shock and Score-Informed group, and is representative of the types of curves found for the Shock and Score, and the Score Only conditions over the 25-min. session. Figure 5 shows the sequential responses for the Shock and Scores-Informed group over the first 5 min. It is representative of the types of curves found for the Shock and Score, and the Score Only-Uninformed groups during the first 5-min. The sequences which are presented are the + + (Score response by *A* followed by Score response by *B*), − − (Shock response by *A* followed by Shock response by *B*), + − (Score response followed by Shock response), and − + (Shock response followed by Score response) sequences. The se-

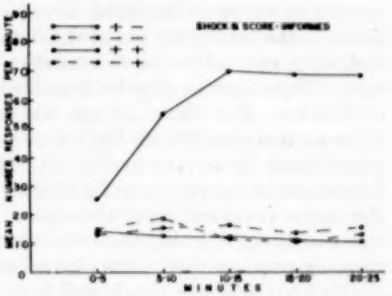

FIG. 4. Mean number sequential responses per minute for the Shock and Score-Informed group over blocks of 5 min. each for the 25-min. session.

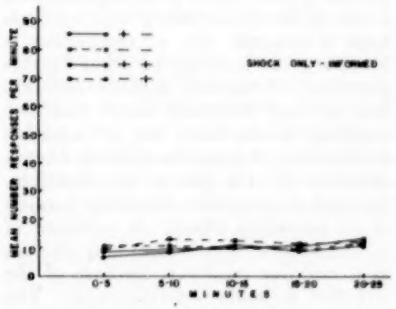

FIG. 3. Mean number sequential responses per minute for the Shock Only-Informed group over blocks of 5 min. each for the 25-min. session.

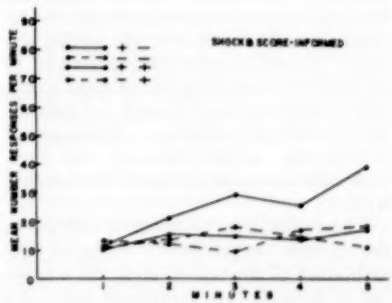

FIG. 5. Mean number sequential responses per minute for the Shock and Score-Informed group over successive 1-min. intervals of the first 5 min.

quence curves were included in order to aid in the interpretation of the data and were not subjected to statistical test. These curves may be described as follows. For those groups which show no learning (Shock Only groups represented by curves in Fig. 3), all four sequence curves maintain roughly the same constant level throughout the course of the experiment. For those groups which show learning (Score Only and the Shock and Score groups represented by curves in Fig. 4 and 5), the + + sequence curve shows large increases, while the other three sequence curves hover at about the same constant level, with possible slight decreases toward the end of the experiment. Therefore, Principle *a* seems to hold; however, Principle *b* is dubious.

Discussion

In setting up the design of the present experiment, two primary questions were raised: (*a*) Would learning take place in such a minimal social situation? (*b*) What would be the effect of Reward-Punishment variables and the factor of Information (Informed, Uninformed) upon results?

First, the results indicate clear evidence of learning, i.e., Score button responses increased even in the first 5 min. of the session. The Shock button responses, however, did not seem to noticeably decrease or increase during the entire session. Second, reward (Score) seemed to be more effective than punishment alone, and a combination of reward and punishment was highly effective. Third, social information added nothing. In fact, such differences as were seen were negative.

Theoretical interpretation is difficult in the present experiment. Detailed examination of the results suggests a number of possible interpretations and rules out others. First of all there are two possible types of learning that might occur: (*a*) Each *S* may simply change his over-all rate of responding under the influence of the scores and shocks received from the other *S*, but without favoring either one of his two responses over the other, or (*b*) *S*s may develop differential response rates to the two buttons. While it is possible that learning of the first type would be largely independent of the actions of each *S*'s partner, it is clear that learning of the second type cannot occur without an interaction between *S*s in each pair which yields differential consequences to the two responses. The second type of learning was the only one found as shown by the failure of the Shock responses to participate in the upward trend of the scores. Therefore, the matter of primary concern is the second type of learning.

Two groups appeared to have little change in behavior or learning. These were the Shock Only groups. It was expected that punishment (Shock) alone would have more of an effect than is shown in the present study. Since the rate of response in the Shock Only conditions remain relatively stable, rather than decreasing in time, one might expect a tendency for *S*s in these groups to learn to favor the non-shock button since a sequence of responses on these buttons would eliminate the punishment. Inspection of the curves, however, shows that this did not occur. One possible explanation of the fact that *S*s did not learn to press the non-shock button significantly more often than the shock button is that the over-all response rate is low for the Shock Only groups, and perhaps a minimal rate of responding is necessary for learning to occur in this situation. A second possible explanation is that receiving shock and not receiving scores both act as negative reinforcers. A sequence of shock-button responses by the pair of *S*s, therefore, may act as a negative reinforcer because of its punishing effect. A sequence of non-shock button responses may also act as a negative reinforcer because of the fact that no score is forthcoming. The final effect is a counterbalancing of negative reinforcers leading to continual oscillation between the buttons.

Now remains the critical problem of

how learning took place in the other four groups. It is necessary to account for a rise in the scores and for the fact that the shock responses remain essentially unchanged. First, the social information factor has been shown to be unimportant in this experiment. Verbal reports of *S*s in the Uninformed groups indicated that none of them were aware of the fact that another *S* was serving in the experiment. In fact, most *S*s under this condition stated that they thought they were working against a set pattern which was being reproduced by an electronic circuit or that they were working against a machine. Many *S*s stated that they had absolutely no control over what scores or shocks they received since the apparatus operated automatically. These verbal reports clearly give little indication of what actually happened or how it came about that these *S*s increased their score responses so nicely. Some explanations based on effects of rewarding and punishing specific responses, however, lead to interesting conclusions. For example, if one examines some of the possible critical sequences of responses the following matrices are obtained:

Type I				*Type II*			
Sequence	*A*	*B*	*A*	Sequence	*A*	*B*	*A*
1	+	+	+	5	+	+	−
2	−	+	−	6	−	+	+
3	+	−	−	7	+	−	+
4	−	−	+	8	−	−	−

In the above matrices, the + specifies the Score button response while the − refers to the Shock response. *A* and *B* refer to the two *S*s, i.e., in Sequence 1, S_A presses a Score button; this is followed by a Score button response by S_B which, in turn, is followed by a Score response by S_A. In the Type I matrix, it would be expected that any one of the sequences of responses made by S_A, then S_B, then S_A would result in an increase in that particular sequence because S_A will tend to repeat responses followed by reward in that particular sequence. In the Type II matrix a decrease would be expected in that particular sequence because S_A will tend to change a response followed by punishment in that particular sequence. If the above assumptions are accepted it is found that, on the average, they will lead to learning because only + + + will give + + + + + sequence; the others give + − + − − + − − sequences. At least two additional assumptions, therefore, must be made: (*a*) some "random oscillation," and (*b*) a dominance or "trapping" characteristic of reward; i.e., + + + sequence must have a greater effect than any minus sequence. This latter assumption is empirically reasonable because Score groups learned better than Shock groups. Turning to the actual sequences, there is some evidence favoring the above interpretation. The only sequence that increases markedly in the groups that learned is the + + sequence.

Summary

The present experiment was designed to investigate the influence of reward (Score) and punishment (Shock) in a minimal social situation. Two *S*s, isolated from each other, were provided with two push buttons by means of which each could give the other shock and scores.

The *S*s, 90 pairs, were divided into two major groups, an Informed and an Uninformed group. All *S*s in the Informed group were told that another *S* was serving in the experiment and that the other *S* controlled his Shocks and Scores. In the Uninformed groups, *S*s were not told that they were in a social situation. The two major groups were subdivided into three Reward-Punishment subgroups in which *S*s could give each other Shock Only, Score Only, or Shock and Score.

The results indicate that learning occurred in the Score Only, and Shock and Score conditions but not in the Shock Only conditions. Learning, therefore, was demonstrated in those groups where *S* had an opportunity to receive a reward (Score). Learning under these conditions occurred regardless of whether or not *S* was told that another *S* was serving in the experiment. There were no significant differences attributable to whether or not *S* knew he was working with another *S*.

It was concluded that the major factors controlling the behavior of *Ss* in the present study were the Reward-Punishment variables, that awareness of the sociality of the situation had little effect, and that reward had a greater effect than punishment.

References

1. Duncan, D. E. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.
2. Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.
3. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
4. Sidowski, J. B., Wyckoff, L. B., & Tabory, L. The influence of reinforcement and punishment in a minimal social situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 115–119.
5. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
6. Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.

(Received September 12, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

EFFECTS OF VARIED DISTANCE ON SHORT-RANGE SIZE JUDGMENTS [1]

NOËL JENKIN [2]

University of New Brunswick

The "law" of phenomenal regression to the real object proposed by Thouless (**9**) predicts that an *S*, matching a standard object at a far distance with a comparison object at a near distance, will choose a value intermediate between the physical size of the standard and one which would subtend the same visual angle as the standard. The generality of this assumption has been challenged by later writers (**1**, **2**, **3**, **7**, **8**) who have clearly shown that such a "law" is true only for a given attitude and understanding of the task, i.e., the "analytic," "retinal," or "stimulus" attitude. Under the contrasting type of condition, where *S* is set to make an "objective" match, i.e., to adjust the variable to a physical size equaling that of the standard, not only is "size constancy" generally achieved, but "overconstancy" is commonly found. This means that in a situation where the standard stimulus is far and the variable comparison stimulus is near, a match will be made which in its physical dimensions exceeds the value of the standard. Furthermore, as shown by the work of Gibson (**2**), Gilinsky (**3**), and Smith (**8**), this effect, within a certain limit, increases with distance, so that a receding object of constant physical size is progressively overestimated.

The investigations which have with fair unanimity established this finding, have mostly used distances of considerable extent. Gibson's farthest distance was about half a mile, Smith's was 320 ft., and Gilinsky's was 4000 ft. All three investigators supplied a visual field rich in cues of distance. Chalmers (**1**) attempted with apparent success to eliminate these secondary cues of distance and found progressive overestimation from binocular judgments at 15 and 20 ft. Overestimation in decreasing amount was then maintained up to 80 ft.

It is evident that overestimation, over a large range of distances is a well-attested phenomenon, and it seems, as Smith (**8**) has pointed out, that this is not readily accounted for in terms of the space-error hypothesis of Holway and Boring (**5**). The generality of the phenomenon requires testing by the study of judgments at distances shorter than 20 ft.

An investigation by the present writer (**6**) suggested that larger estimates would be obtained with the comparison object at 1 or 2 ft. than with the comparison at 15 ft., when the standard object remained constant at 30 ft. The purpose of the present study was to check this finding under properly controlled conditions, and to test the hypothesis, also suggested by the former investigation, that accuracy of judgment would be greater when the comparison was further from *S* than when it was at the 2-ft. distance.

METHOD

Apparatus.—The apparatus consisted of white cardboard squares. The standard card, 4 in. square, was mounted on a wooden post, 47½ in. above floor level, and was positioned directly in front of *S*, at a distance of 20 ft. from his eyes.

[1] The author is indebted to Dr. Ray Hyman of Harvard University for critically reading the first draft of the present manuscript.

[2] Now at The Training School at Vineland, N. J.

The comparison series was a set of squares ranging from 2¾ in. to 5¾ in. by steps of ½ in. They were individually displayed upon a platform at the same height as the standard, but positioned on *S*'s left at an angle of 90° with respect to the direction of the standard. The platform measured 27 × 24 in. and was papered in a dull green. A background of similar material, measuring 15 × 24 in. was fastened at the rear of the platform. Concealed pins and a slight depression, invisible to *S*, served to maintain the cards in vertical position in the middle of the platform. No special background was provided for the standard object. A large room, well and constantly lighted by daylight and containing no interposed objects, provided a setting for the apparatus.

Subjects.—The *S*s were 48 students in elementary psychology courses at the University of New Brunswick Summer School, 24 of each sex, who were required to serve in the experiment as part of their course work and who received credit points for doing so. At the time tested, all *S*s were naive with regard to size and distance theory and experimentation, and all had approximately normal vision or vision satisfactorily corrected. They were randomly assigned to four subgroups of 12 *S*s each, with an equal number of males and females in each group.

Design.—Half of the *S*s (Group I) first made judgments of the standard when the stimuli were equidistant at 10 ft. They then made their judgments at the disparate distances. With the other half of the sample (Group II), this procedure was reversed, so that the equidistant judgments were made last. Half of the *S*s in each main group were presented with the comparison first at 10 ft. and then at 2 ft. (Sequence A). The other half of each group was treated with the converse order (Sequence B). All four subgroups therefore made judgments with standard and variable displayed at (*a*) 10 ft. and 10 ft., (*b*) 20 ft. and 10 ft., (*c*) 20 ft. and 2 ft. For Group IA the order was *a*, *b*, *c*; for Group IB it was *a*, *c*, *b*; for Group IIA it was *b*, *c*, *a*; and for Group IIB it was *c*, *b*, *a*.

Procedure.—The height of each seated *S*'s eye level was adjusted to correspond with that of the stimulus objects by providing him with an appropriate cushion or cushions. Instructions to *S* emphasized an objective attitude. Each card as it was presented on the platform was to be judged as larger, equal to, or smaller in physical size, with respect to the standard. When a card was described as equal, it was to be *S*'s opinion that its dimensions would measure the same as those of the standard card. From each *S*, two judgments were secured at each distance, one from an ascending series and one from a descending series. For each subgroup, half of the *S*s started with an ascending series and half started with a descending series. Prior to the first series one judgment was secured in order to illustrate the procedure. The card used for this practice trial was one outside the range of the comparison series proper. For *S*s beginning with an ascending trial, the practice card was 6 in. and for those beginning with a descending trial, it was 2½ in. The *S* was required to watch the standard, while *E* placed the comparison card in position. The *S* then turned his head and observed the latter. He was not prevented from looking back and forth if he wished to do so. After making his judgment, he was required to watch the standard again while *E* replaced the comparison card. After the judgments were obtained at the first distance, the stimulus objects were quickly moved to the second position. Each card in the comparison series was seen by the *S* no more than twice. A second, identical set of cards was used for the third judgment. Each series was begun with a different sized card.

Results

The mean size judgments for each subgroup at each distance are shown in Table 1. They are expressed in raw units in order to facilitate comparison with the physical size of the standard being judged, i.e., 4 in. A preliminary analysis of variance of all judgments at the disparate distances showed that Sex Differences and interactions of Sex with Groups (prior versus latter exposure to the equidis-

TABLE 1

Mean Size Judgments of 4-in. Standard at Two Distances

Distances	Group I (10 ft./10 ft. Judgments First)		Group II (10 ft./10 ft. Judgments Last)		Grand Means
	Seq. A*	Seq. B*	Seq. A	Seq. B	
20 ft./10 ft.	4.006	3.980	3.980	4.095	4.014
20 ft./2 ft.	4.231	4.069	4.184	4.345	4.205
10 ft./10 ft.	4.023		4.038		4.031

* In Sequence A, the comparison stimulus occurred first at 10 ft. and then at 2 ft.; in Sequence B the reverse was true.

TABLE 2

ANALYSIS OF VARIANCE OF SIZE JUDGMENTS

Source	*df*	*MS*	*F*
Groups (G)	1	1496	0.79
Sequence (S)	1	157	0.08
Subjects	46	1891	
Distances (D)	1	9184	25.87**
G × D	1	291	0.82
S × D	1	91	0.26
G × S	1	456	0.24
G × D × S	1	3093	8.71*
Residual	42	355	

Note.—The Groups variable represents prior exposure to equidistant condition vs. no prior exposure.
* $P < .01$.
** $P < .001$.

tant condition), with Sequence, and with Distances were all nonsignificant ($P > .05$) hence the data from both sexes were combined in subsequent analysis. Examination of the combined data for the 20 ft./10 ft. and the 20 ft./2 ft. distances showed that the means tended to be proportionate to the *SD*'s. The variances were therefore stabilized by means of a logarithmic transformation and an analysis of variance was performed on these measures. Table 2 gives the results of this analysis, which strongly supports the hypothesis that judgments at the 2-ft. distance would be significantly larger than those at the 15-ft. distance. It may also be noted from Table 2 that there is a negligible difference between groups with prior versus latter exposure to the equidistant condition. A nonsignificant *F* value is also found for Sequence (order of presentation of the two disparate distances) and for each of the simple interactions. A significant triple interaction indicates that prior exposure to the equidistant stimuli has differential effects on judgments at the disparate distances, depending on which distance is experienced first. The variance estimate for the triple interaction is not so large as to render equivocal the interpretation of a difference between means for the two distances. It is very clear that the distance variable accounts for the great majority of variance in the data.

The equidistant judgments served as a control in testing the hypothesis that the accuracy of judgment would be greater at the 10-ft. than at the 2-ft. distance. Since the group making equidistant judgments first did not differ significantly from that making them last ($t = .25$), the data from the 10 ft./10 ft. distance were combined. A test of the hypothesis that the mean of these judgments did not differ significantly from a parent population whose true mean was 4.0 in. (the physical dimension of the standard) found a t of 1.06, $P > .25$. A similar test applied to the judgments made at 20 ft./10 ft. found a t of .15. Therefore it was concluded that there was no reason to suppose that under either of these conditions the sample was deviating by more than chance fluctuation from a value equal to the objective, physical size of the standard being judged. The same test applied to the 20 ft./2 ft. judgments found a t of 3.72, $P < .001$. At this distance, the sample mean differs significantly from an assumed parameter of 4 in., in the direction of overestimation. It would appear to be unlikely that a "space error" is responsible for this overestimation, in view of the nonsignificant difference from 4 in. of the equidistant judgments.

DISCUSSION

The distant object remained constant throughout the experiment. Since it was matched from a comparison series both at 2 ft. and 10 ft., it follows that the two matching cards were also subjectively equal to each other. Actually, the one chosen at the near position tended to be physically larger than that chosen at the

farther position. At first thought it might seem that this trend is consistent with the geometry of perspective, i.e., the object decreasing in size with distance. Further reflection, however, makes it clear that an object of given physical size appears *smaller* when it is close, and therefore a physically larger match is required to establish subjective equality with the standard. Apparent size thus significantly increases over the short distance interval from 2 ft. to 10 ft.

Furthermore, the average judgment at the near position exceeded the physical size of the standard and at the far position was exceedingly close to the physical size of the standard. It is therefore clear that in these judgments there was no "compromise" between retinal size and objective physical size. On the contrary, in fact, at one of the distances, "overconstancy" was reliably demonstrated.

These findings complement those of Chalmers (**1**), Gibson (**2**), Gilinsky (**3**) and Smith (**8**). The increase of apparent size with distance found by these writers under long-range conditions is now seen to take place at very short distances also. We may therefore offer a tentative generalization by saying that objective judgments in the size-distance situation appear to follow a trend *opposite* to the direction predicted by the rules of geometrical perspective.

A further possibility is suggested by the present demonstration that a standard is significantly overestimated when the comparison is very close to *S*. It appears probable that in judging the size of an object at a given distance by means of a comparison series, there may be for the latter a point or range of optimum location for greatest accuracy. The limits of this function, at different distances of the standard, is a matter which requires further exploration.

Hitherto, overestimation of size has been explained—e.g., by Smith (**8**)—by error in the estimation of distance. If distance is overestimated, it is reasoned, size will be likewise overestimated. Progressive overestimation of size with increasing distance (**2, 3, 8**) lends some weight to this hypothesis. Such a view, while not inconsistent with the data of the present experiment, provides a less satisfactory basis for explaining short-range overestimation of size. Hermans (**4**) has shown that a decrease in perceived size is coupled with an increase in binocular convergence. This finding, also consistent with the present results, may furnish a more useful point of departure for further study of the phenomenon.

Summary

The present experiment studied the trend in size judgments under objective instructions at two short viewing distances. Judgments of a standard 20 ft. away were obtained with the comparison object presented at two distances, from four groups of 12 *S*s each. Two groups were first required to give judgments with the stimuli equidistant at 10 ft. The other two groups gave judgments at the disparate distances, and then made equidistant judgments. Two groups made judgments with the sequence being 20 ft./10 ft. and then 20 ft./2 ft. The other two groups had this order reversed. Analysis of variance showed that prior experience with the equidistant stimuli had no significant effect on the subsequent judgments and did not simply interact with either of the other variables. A highly significant difference between the two distances, relatively independent of the effects of sequence and prior experience, confirmed the prediction that the nearer object would yield the larger match, and also that accuracy would deteriorate from the far to the near position of the comparison.

References

1. Chalmers, E. L. Monocular and binocular cues in the perception of size and distance. *Amer. J. Psychol.*, 1952, **65**, 415–423.
2. Gibson, J. J. *Motion picture testing and research*. AAF Aviation Psychology Program Res. Rep. No. 7. Washington, D. C.: U. S. Govt. Print. Off., 1947.
3. Gilinsky, A. S. The effect of attitude upon the perception of size. *Amer. J. Psychol.*, 1955, **68**, 173–192.
4. Hermans, T. G. The relationship of convergence and elevation changes to judg-

ments of size. *J. exp. Psychol.*, 1954, **48,** 204–208.

5. Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54,** 21–37.
6. Jenkin, N. Size-distance judgment as a function of personality, intellect, attitude and stimulus variables. Unpublished doctor's thesis, Harvard Univer., 1957.
7. Joynson, R. B. The problem of size and distance. *Quart. J. exp. Psychol.*, 1949, **1,** 119–135.
8. Smith, W. M. A methodological study of size-distance perception. *J. Psychol.*, 1953, **35,** 143–153.
9. Thouless, R. H. Phenomenal regression to the real object. *Brit. J. Psychol.*, 1931, **21,** 339–359; 1932, **22,** 1–30.

(Received October 3, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

THE INTERACTION OF ASSOCIATION VALUE AND STIMULUS CONFIGURATION IN SIZE ESTIMATION

MARGARET E. DOW
University of Wisconsin

AND

JESSE E. GORDON
Montana State University[1]

The study of possible relationships between emotional factors, value structures, and personality "needs," and various parameters of the perceptual process has enjoyed some popularity in recent years. An aspect of perception which has received much attention in these studies is that of size estimation, stimulated largely by Bruner and Goodman's (1) study in which size estimations of coins were found to be related to the economic background of the children making the judgments. That factors other than the value of the coins may play an important role in producing these results is suggested by an experiment by Bruner and Rodrigues (3), who varied the material of coin-like stimuli, and Bruner and Postman (2), who varied the designs on the faces of the coin-like stimuli. Evidence from these studies suggests that variations in these structural properties of the stimuli also contribute to differences in perception, either in addition to or regardless of the associational value of the stimuli. But Bruner and Postman's study is ambiguous in this regard, since the designs they used were representational of a dollar sign, a swastika, and a neutral symbol. Thus associational value was confounded with structural (design) differences of the stimuli.

A study by Singer (4) was the immediate stimulus for the present study. Singer found that by slightly changing a circular stimulus design so that it could schematically represent a smiling and/or a frowning face, he could produce greater size estimations than when the stimulus circle contained a nonsensical design, or was blank. Singer found no differences between size estimations of the frowning and the happy faces. He concluded that meaningfulness (or association value, as it will be referred to hereinafter) is a significant variable in size perception. However, it should be noted that in Singer's experiment, association value and structural properties of the stimuli are not independent of each other. One might interpret his results as simply indicating that two of his configurations of lines, those which he called happy and frowning faces, are perceived as being larger than the other stimuli as a result of the manner in which the lines of the design curve, creating perceptual phenomena not unlike the Müller-Lyer illusion, etc., and that these larger size estimations are completely independent of the associational value of the stimuli.

The present experiment was designed to test the same major hypothesis as did Singer—namely, that there are differences in the size estimations of visual stimuli as a function of the association value of the stimuli and of the pleasant or unpleasant "tone" of the associations. However, an attempt has been made in this experiment to vary this association value independently of the stimulus configuration by associating pleasant concepts, unpleasant concepts, and

[1] This research was carried out while the junior author was a U. S. Public Health Service post-doctoral fellow at the University of Wisconsin.

nonsense syllables to the same standard stimuli. The results of such an experiment might also shed more light on the question of whether the differences in size estimations of the various stimuli studied in the experiments reviewed above were due to autochthonous qualities of the perceptual process itself, or may be ascribed to other functional components of behavior which are called into play after the perceptual act has been committed, such as those involved in reproduction tasks. The present experiment has been designed to eliminate or control these latter variables to a great extent, through the use of covariance analysis in which general response dispositions in making size estimations are set aside.

It was also hypothesized that there would be an interaction between association value and configurational aspects of the stimulus figures. Specifically, it was expected that figures consisting of curved lines should heighten any effects produced by pleasant association value, and that angular figures should intensify the effects of unpleasant associations. These notions are derived from assumptions commonly made by artists and architects regarding the effects on the viewer of the curved or angular characteristics of shapes and lines.

Method

The *S*'s were 40 undergraduate students at the University of Wisconsin. Each *S* was asked to select the five words from the Kent-Rosanoff word-association list which had the most pleasant meanings for him, and five words which were the most unpleasant. Following this, *S* was asked to make five practice control settings of the size estimation apparatus, using a 5 × 5-in. blank white card on a black ground as the stimulus. The size estimation apparatus consisted of a large black L-shaped board mounted at desk height before the seated *S*. Size estimations were made by *S* by decreasing the size of a white square directly behind the L-shaped board with its face contiguous with the back surface of the board. This was done by having *S* release a positioning knob attached to the white square, which allowed the square to slide downward at a 45° angle until the portion of the square still visible within the angle was of the desired size, at which point *S* could turn the knob to hold the sliding square in place. The *S* was also permitted to move the square up, if he felt he had let the slide down too far. The *E* could then read the size of the setting from a centimeter scale placed on the back of the sliding white square. Settings could range from 0 to 30 cm. The *S* was told to "make the size of the slide smaller until it seems to be the same size as the square hanging there (the stimulus)." The *E* changed the stimulus square for each practice setting, but the five stimuli were in actuality identical in size, color, and position.

Following these practice, control settings, *S*s were randomly assigned to the eight experimental groups: Curved Pleasant Association, Curved Unpleasant Association, Curved Nonsense Association, and Curved No Association; and Angular Pleasant Association, Angular Unpleasant Association, Angular Nonsense Association, and Angular No Association. Each *S* in the Pleasant Association groups learned to associate one of the Kent-Rosanoff words which he had selected as pleasant to each of the five stimulus figures. The *S*s in the Unpleasant Association groups learned to associate the words they had selected as unpleasant to the stimulus figures. The Nonsense Association groups learned to associate five nonsense syllables to the five stimuli, and the No Association groups were simply shown the stimulus figures without associations. The method of learning the associations in the various association groups was to present each stimulus design to *S*, with its associated word clipped to the back of the stimulus card. After the first trial, in which each card was shown for 3 sec. and then turned over so that its associated concept or nonsense syllable was seen by *S*, *S* was asked to identify the concept or nonsense syllable associated with each card as the figure was presented. Each stimulus card was shown for 3 sec. and, if a correct response was not given in that time, then turned over to reveal the associated word or syllable to *S*. The stimulus cards with the designs were shown in random order for each trial. The *S*s learned to a criterion of errorless anticipations in three consecutive trials of the set of five designs. The *S*s in the No Association group were shown the five designs for 3 sec. each, for a total of 15 times, which compares with a mean number of trials in the other three groups of 11.4.

The stimuli used for size estimations consisted of five basic designs derived from the Bender-

Gestalt test, and modified so that each design consisted only of straight lines. Another set of stimuli were then prepared in which the general configurations of the first set were maintained except that the designs for the second set consisted entirely of curved lines. For example, a square in the angular set was reproduced as a circle for the curved-line set. The designs were drawn in black ink on white 5 × 5 in. cards, and covered an area of about 3 × 3 in. on the cards. The stimuli used for the associative learning trials were 3½ × 3½-in. reproductions of these cards.

The stimuli for the size estimation trials were hung on a black background at a distance of 20 ft. from *S*, in a darkened hallway. All *S*s were tested at about the same hours in the evenings in an effort to hold visual cues constant.

The experimental trials consisted of 25 settings of the size estimation apparatus by each *S*, the 25 settings being divided into five massed trials, each of which consisted of a different order of presentation of the five designs.

Results

Each *S*'s scores for the data analyses were the sums of his size estimation settings in each block of five trials. Thus each score represented five settings. An over-all analysis of covariance of these data indicates that no significant differences among association groups, or between types of designs, were found. Because of the possibility that any effects of association value on size estimation might be subject to extinction as a function of amount of exposure to the stimuli, the data were analyzed using only the first five experimental settings by each *S* (one setting for each design). In this analysis, a significant interaction was found. Table 1 reports this analysis.

TABLE 1

Analysis of Variance of Errors of Estimate, All Conditions, First Five Trials ($N = 40$)

Source	*df*	*MS*	*F*
Among Association Conditions	3	28.54	.71
Angular-Curved	1	6.07	.15
Interaction	3	118.15	2.93*
Within Samples	31	40.33	

* Significant at .05 level.

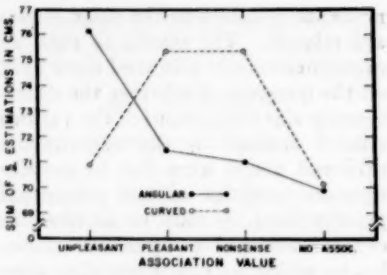

Fig. 1. Mean sums of first five size estimations.

Figure 1 presents the adjusted means of the sums of each *S*'s first five judgments. The adjusted means of the curved and angular design groups are similar only in the No Association condition.

As a more specific test of the hypothesis that there is an interaction between the hedonic "tone" of the association value of the stimuli and the angularity or curvilinearity of the stimulus configurations, two more *S*s were added to each of the Pleasant Association groups and to each of the Unpleasant Association groups, and the data of these four groups reanalyzed, excluding the Nonsense Association and No Association groups. The results of an analysis of covariance on all the settings made by these Pleasant and Unpleasant Association groups ($N = 28$) did not reveal the hypothesized interaction. The results of an analysis of the first five settings only of *S*s in these four groups also failed to produce a significant interaction. These results indicate that the significant interaction reported in Table 2 cannot be accounted for by the interaction of curvilinearity-angularity with only

Pleasant Association and Unpleasant Association.

Discussion

The results of this experiment indicate that association value per se does not influence size estimations. Only in interaction with configurational characteristics of the stimuli do associations affect the perception of size. Further, the results indicate that these effects are quite unstable, disappearing quite readily with increased exposure to the stimuli.

The interaction of Pleasant and Unpleasant Association with Angularity and Curvilinearity is in the predicted direction. However, this interaction is not statistically significant alone; in this experiment, it was also dependent on the interaction with Nonsense Association. The results suggest that there is some validity for identifying angularity with unpleasantness, and curvilinearity with pleasantness; the interaction produces an illusion of larger size. However, the finding that Nonsense Association interacts with the stimulus variables studied in this experiment in the same manner as does Pleasant Association is not readily explainable. There is the possibility that the learning of the associations provided a rewarding, or pleasant, experience for the subjects in the Nonsense and Pleasant Association conditions, whereas the learning experience in the Unpleasant Association condition served to emphasize the unpleasantness. This might account for the similarity of response to Pleasant and Nonsense Associations.

Summary

An experiment was designed to test the hypothesis that the association value and the pleasantness or unpleasantness of the association to stimuli affects the perception of the size of the stimuli. In addition, it was hypothesized that curvilinearity or angularity, as stimulus configurational variables, interact with association value such that curved figures would enhance the effects of pleasant associations and angular figures would enhance the effects of unpleasant associations. Forty undergraduate *Ss* each made five size estimations by varying a stimulus until it matched the standard to provide measure on a matching variable. Concepts selected by each *S* as pleasant or unpleasant were then equated with either curved-line or angular-line designs which served as the stimuli. A third group had nonsense syllables associated with the figures, and a fourth group had no associations. The experimental trials consisted of matchings of the variable stimulus to each of five design stimuli. The results indicate no significant differences among the groups in their size estimations of pleasant, unpleasant, nonsense, or no association figures, or of curved or angular figures. A significant interaction between curved-angular figures and the association variables was found only for *Ss'* first experimental judgments of each stimulus. The results are interpreted to indicate that association value has an unstable effect on size estimations, depending on configurational aspects of the stimuli, and on relatively few exposures to the stimuli.

References

1. Bruner, J. S., & Goodman, G. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42,** 33–44.
2. Bruner, J. S., & Postman, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27,** 203–208.
3. Bruner, J. S., & Rodrigues, J. S. Some determinants of apparent size. *J. abnorm. soc. Psychol.*, 1953, **48,** 17–24.
4. Singer, J. L. Personal and environmental determinants of perception in a size constancy experiment. *J. exp. Psychol.*, 1952, **43,** 420–427.

(Received October 22, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

SEMANTIC GENERALIZATION WITH EXPERIMENTALLY INDUCED ASSOCIATIONS

WENDELL E. JEFFREY AND RICHARD J. KAPLAN

University of California, Los Angeles

One approach to the study of meaning has come under the rubric of semantic generalization. Such generalization is usually indicated by positive transfer presumed to result from a similarity defined in terms of an association between either two signs, two objects, or a sign and an object. Traditionally, the strict mediated generalization study, following an A—B, B—C, A—C, paradigm has been excluded as not being "semantic in character" (**2**). This leaves situations mainly of the sort that may be represented by the paradigms, A—B, A—C, B—C, and A—C, B—C, A—B, facilitation being expected in the third task as the result of the other relationships specified.

Foley and Cofer (**2**) demonstrated semantic generalization in a situation where the recall of a test list of words was facilitated when it followed training on a serial list made up of synonyms and homonyms of the test list. Osgood (**4**), however, points out that because of the constant relation between the training and test lists, the *Ss* may have reconstructed, rather than recalled, the test list once the relationship was recognized.

Peters (**5**) performed several experiments following the A—B, C—B, A—C paradigm without obtaining particularly impressive results, except in one experiment where meaningful words that bore a part-whole relationship to each other were used. It appeared, however, that Peters' learning criterion was not sufficiently high if a single anticipation trial rather than learning were to be used as the measure of facilitation.

Nevertheless, the Peters studies suggest that semantic generalization could be more adequately studied if the essential relationships were brought under experimental control by using nonsense syllables. The semantic relationships that are presumed important can then be created in the experimental situation.

An experiment by Bugelski and Scharlock (**1**) on mediated generalization provided a convenient and efficient design for the type of study proposed. However, rather than use only an arrangement of syllables that would presumably produce interference with which to contrast the facilitation experience, a third subset of syllables was added for comparison with the effects of both interference and facilitation subsets.

Thus, the following experiment was designed to (*a*) demonstrate semantic generalization when the "semantic relationships" were under experimental control, and (*b*) to evaluate semantic generalization through both transfer and interference effects.

METHOD

Apparatus and procedure.—Each *S* learned three lists of 12 pairs of nonsense syllables by the anticipation method. These lists were learned in three sessions separated by 48-hr. intervals. Each list was made up of three subsets of syllables, two experimental and one control, so that each *S* provided both experimental and control data. The E_1 subset was four syllable pairs which followed the paradigm, A—B for Session 1, A—C for Session 2, and B—C for Session 3. The second experimental subset of

syllables (E_2) followed the same daily sequence as did the E_1 syllables except in Session 3 when the syllables of the E_2 subset were rearranged so that the two syllables making up each pair had not been common responses to a single stimulus as was the case for the E_1 syllables in this session. The control subset (C) followed the paradigm A—B, D—C, B—C, and as can be seen, differed from the experimental groups in Session 2 in that the C syllables were not associated with A syllables but with a new set of stimuli. The three subsets of syllables were combined in an unpredictable order in each list.

In order to balance the difficulty level of various syllables in the various subsets, *Ss* were divided into three groups and lists were prepared such that each set of syllable pairs was used in each of the experimental as well as in the control subset. Thus there were nine lists of 12 syllable pairs, one for each group for each of three sessions. The syllables were taken from Glaze's lists (**3**) with association values between 80% and 87%.

The instructions were the usual ones for paired-associate learning with the exception that *Ss* were requested to pronounce the syllables rather than spell them. The syllables were presented with a Gerbrands M-1 memory drum and *S* was separated from *E* by a hardboard screen. A 3-sec. interval separated the presentation of the stimulus and the stimulus-response pair, and another 3 sec. elapsed before presentation of the next stimulus. A 6-sec. interval separated repetitions of the list. A 1-min. rest was introduced after each nine repetitions of the list unless *S* was in the process of meeting the learning criterion. All lists were learned to a criterion of two successive correct repetitions followed by three more repetitions of the list.

Subjects.—The *Ss* were 21 male students drawn from psychology classes. None had had previous experience with nonsense syllables.

Results and Discussion

The mean number of errors to criterion for each group, for each subset, is presented in Table 1. The subset E_1, that was expected to indicate semantic generalization through fewer errors in Session 3 than the other subsets, provided results in the predicted direction. The interference subset (E_2), however, did not produce a mean that differed from that obtained for the control (C) subset. The analysis of variance summarized in Table 2 indicates a significant Treatment effect ($P < .01$). The group differences, *S* differences, and interaction did not produce significant *F* ratios. Thus, there are no limitations placed on the interpretation of the obtained difference from the statistical standpoint. However, if one tries to interpret these data within the usual associationistic interpretation of mediation phenomena, there are theoretical problems. The reasoning necessary to the prediction of facilitation for E_1, also predicts interference for E_2, unless some auxiliary process is proposed, or unless one can find some reason for rejecting the present control group as inadequate.

There are several ways in which this latter possibility may be evaluated. Even though the analysis of variance of errors in Session 3 showed no group differences or significant interaction with

TABLE 1

Mean Errors to Criterion for the Different Groups and Syllable Subsets

Groups	Syllable Subsets		
	E_1	E_2	C
1	19.8	21.6	25.4
2	19.7	25.6	29.3
3	23.1	34.6	28.3
Total	20.9	27.2	27.7

TABLE 2

Analysis of Variance of Errors to Criterion in Session 3

Source	*df*	*MS*	*F*
Between groups	2	216.0	.50
Error	18	431.6	
Between treatments	2	301.0	7.68*
T × G	4	74.2	1.89
Error	36	39.2	
Total	62		

* $P = .01$.

other variables, the groups were compared in Session 1. Mean trials to criterion for Session 1 were 20.0, 20.4, and 21.2 with standard errors of 1.86, 3.47, and 1.67, respectively. Thus, there is no evidence of differential learning ability among the three groups when the learning conditions were identical. In Session 2, the control condition can be differentiated from the two experimental conditions, inasmuch as it involved associating the C syllables with a new set of stimulus syllables rather than with the previously used A syllables. The mean errors to criterion per syllable pair were 8.95 and 9.47 for the experimental conditions and 7.77 for the control conditions. The appropriate analysis of variance and *t* tests indicate that there are significantly fewer errors on the control syllables than on the experimental syllables. This difference is not surprising inasmuch as some interference would be expected in the experimental subsets in Session 2. Whether this difference has any implications with respect to the results obtained in Session 3, however, is another question. Considering the relative magnitude of these means it does not seem plausible to suggest that practice effects, warm-up, or familiarization, were not equalized, and thus this difference, in spite of its significance, does not provide a suitable explanation for the lack of difference between the control and interference subsets in Session 3.

The positive results of this experiment when compared with the essentially negative results obtained by Peters are probably not due so much to the different paradigms involved, as they are to the higher learning criterion, the greater efficiency of the experimental design, and the fact that learning rather than a single anticipation was used as a criterion. Further research will be necessary to establish whether interference can be produced by the semantic generalization paradigm. Interrogation of *S*s gave no indication of awareness of the possibility for mediated transfer in Session 3.

Summary

Previous studies of semantic generalization have used associations provided in the uncontrolled experience of *S*. The present experiment used one possible paradigm of semantic generalization, and with nonsense syllables, provided the appropriate associations through training. With each *S* serving as his own control, that arrangement of syllables that was expected to facilitate performance did so to an extent differing significantly from a presumably neutral arrangement, and an arrangement expected to produce interference. The latter two training procedures, however, did not produce differential learning on the critical task.

References

1. Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44,** 334–338.
2. Foley, J. P., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental studies of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32,** 168–175.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–269.
4. Osgood, C. E. *Method and theory in experimental psychology.* London: Oxford Univer. Press, 1953.
5. Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18,** 20–48.

(Received October 25, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

THE IMMEDIATE EFFECTS OF EMOTIONALITY UPON BEHAVIOR STRENGTH[1]

NORMAN R. ELLIS[2]

Louisiana State University

Hull (**9**) postulated that behavior strength is a multiplicative function of a learning factor, habit strength ($_sH_R$), and a motivational factor, generalized drive (D). $_sH_R$ is defined by changes in behavior due to practice; D is defined jointly by greater energization of behavior following such antecedent conditions as deprivation of food and noxious stimulation. Hull then hypothesized that the total effective drive strength (D) operative at a given moment consists of the relevant drive strength ($\bar{D}$) plus the irrelevant drive strength ($\dot{D}$) derived from all needs coexisting in the organism. $\bar{D}$ is contributed by the need which is reduced by the criterion behavior, whereas $\dot{D}$ is contributed by needs which are historically unrelated to the criterion behavior and whose existence is defined by independent operations.

Hull adduces evidence to support the drive summation hypothesis from studies by Perin (**13**), Williams (**20**), Zener and McCurdy (**22**), Finch (**8**), and Skinner (**17**). Following these early studies, which dealt indirectly with the problem, experiments were specifically designed to test this hypothesis. These studies divide themselves into those concerned with *consummatory* behavior and those dealing with *preparatory* behavior.

Siegel and Siegel (**15**), Amsel and Maltzman (**3**), Siegel and Brantley (**16**), and Wohl (**21**) report positive evidence for drive summation. In each of these studies emotionality, induced by faradic stimulation, was employed as an irrelevant need and either food or water deprivation defined the relevant need. Dry food consumption or water consumption was the criterion behavior. Ellis, Siegel, and Wohl (**7**), using a less intense relevant need—hunger—found negative evidence for the hypothesis.

Several studies have been concerned with the effects of $\dot{D}$ upon preparatory behavior. Experiments by Koch and Daniel (**11**), Siegel (**14**), and Strange (**18**) yield data which do not support Hull's hypothesis. Studies by Kendler (**10**) and Amsel (**1**) report both negative and positive evidence, depending upon the need state utilized and the intensity of the need state. Webb (**19**) and Brandauer (**5**) report experiments which yield positive findings only, for both drive summation and substitution.

Miller (**12**) has reported two studies which suggest that an irrelevant need may function as a relevant need in instances where the relevant need has been eliminated.

The $\dot{D}$ used in the present experiment is called emotionality, viewed as a perseverative response following pain and possessing motivational properties. $\bar{D}$ is generated by food privation. The criterion behavior selected is start-box latency and running time in a linear, alley maze. Each drive condition has been tested at zero, weak, and strong levels yielding unique drive conditions for nine independent groups. The occasion for drive summation and/or substitution has been provided under conditions which appear to constitute a fair test of the hypothesis. The conditions have been outlined by Siegel and Brantley as follows: "(*a*) the drive level is not already maximal, (*b*) the appropriate stimulus occasion

[1] This article is a part of a Ph.D. dissertation submitted to the Graduate Faculty of the Louisiana State University. The author is indebted to Dr. Clyde E. Noble under whose direction the research was performed.

[2] Now at the State Colony and Training School, Pineville, Louisiana.

for the elicitation of the response is presented, and (*c*) stimuli eliciting competing responses are absent" (**16**, p. 304).

Method

Subjects.—The *S*s were 126 male and female albino rats of the Harlan strain. Their ages ranged from 80 to 90 days at the beginning of the experiment, and their weights from 122 to 258 gm., with a mean of 187.3 gm. The *S*s were housed in groups of four in cages of approximately 20 × 10 × 10 in. and were maintained on Purina Dog Chow Checkers and water *ad lib.*

Apparatus.—Two linear mazes and a shock box constituted the apparatus. The maze utilized for pretraining had a start box and a goal box of 18 × 11 × 10 in. inside and a 30-in. straight alley, the inside dimensions of which were 4 × 4 in. The start box was painted flat white; the alley and the goal box, flat black. Both the start box and the goal box had guillotine type doors which, when opened, activated microswitches controlling two electric clocks. These clocks measured start-box latency and running time to the nearest .01 sec. The training-test maze was identical to the pretraining maze except that the alley was 17 ft. 6 in.

The shock apparatus consisted of a 20 × 15 × 14 in. compartment, the walls of which were painted with black and white diagonal stripes. The top was plate glass. The grid floor was made of brass rods $\frac{1}{8}$ in. in diameter and spaced approximately $\frac{1}{2}$ in. apart. A shock current was provided by a constant D.C. source with a rated output of 805 v. A 1-megohm resistor was placed in series in the circuit in the instance of "strong" shock to control current intensity and to minimize changes in *S*'s resistance. For the "weak" shock a 2 megohm resistor was used. A Hunter-Brown electronic timer set at .5 sec. controlled the shock duration.

Procedure.—The experiment was conducted in two sessions, each lasting 44 days and consisting of pretraining, training, and a test. The pretraining regimen was designed to establish specific eating habits, adapt *S*s to handling, and familiarize them with the maze environment. For the first 14 days *S*s were fed wet Purina mash and handled during a 1-hr. period each day. Water was available on the handling tables as well as in the home cage. Following this phase of pretraining, they were given one trial daily for eight days in the short maze with wet mash available for 45 sec. in the goal box. During this phase the 23-hr. food privation regimen was maintained.

At the termination of this 22-day period the training proper was begun. Training was identical to the latter phase of pretraining except that the long maze was used and the training was continued for 21 days. Two measures were recorded for each training trial: latency and running time. Latency was defined as the time interval commencing 10 sec. after *S* was placed in the start box and ending when the start box closed, i.e., when *S* was in the alley and just clear of the door. The *S*s were placed in the start box facing opposite the door. Running time was defined as the interval between the closing of the start-box door and the closing of the goal-box door after *S* had entered.

The criterion test was conducted on the day immediately following the last training day. At this time the 126 *S*s were equally and randomly assigned to the nine cells of a 3 × 3 factorial design. No sex bias resulted from this procedure. One dimension of this design consisted of 0, 12, and 23 hr. food privation, $\bar{D}$. The other represented zero, weak, and strong shock, $\dot{D}$. Thus the nine groups will henceforth be designated 0-n (none), 0-w (weak), 0-s (strong), 12-n, 12-w, 12-s, 23-n, 23-w, and 23-s where the numerals refer to the intensity of $\bar{D}$ and the letters to $\dot{D}$.

Groups 23-n, 23-w, and 23-s were tested under the hunger need that prevailed during training. Groups 12-n, 12-w, and 12-s were allowed to eat for 1 hr. exactly 12 hr. prior to testing. An intensive effort was made to insure that groups 0-n, 0-w, and 0-s were satiated at the time of testing. This entailed allowing *S*s access to wet mash for 2½ hr. or until they had stopped eating for 5 min., whichever time was longer, immediately before testing. Qualitatively, all *S*s in these groups appeared lethargic and fairly quiescent.

The treatments designed to induce emotionality were intermittent faradic stimulation delivered immediately prior to the criterion maze run. Groups receiving no shock were simply exposed to the shock box for 20-sec. Those receiving weak shock were given 10 shocks of .35 mamp. during a 20-sec. interval and *S*s getting strong shock, 10 shocks at .67 mamp. during the 20 sec.

Several precautions were taken to prevent generalization from stimuli associated with the shock situation to stimuli in the test environment. Some of these were: (*a*) the shock procedure was carried out by an *E* who had no other contact with the *S*s, (*b*) the *E* who administered the shock wore black rubber gloves, and (*c*) *S*s were shocked in an environment other than the test situation. The exact procedure consisted of the shock-*E* taking *S* from his cage and placing it in the shock box. Then it was handed to the test-*E* and immediately placed in the start box. Four to 8 sec. elapsed between

TABLE 1

LATENCY ON TEST TRIAL

$\bar{D}$ Food Deprivation	$\dot{D}$								
	No Shock (n)			Weak Shock (w)			Strong Shock (s)		
	Mean	*SD*	Mdn.	Mean	*SD*	Mdn.	Mean	*SD*	Mdn.
0 hr.	3.7	3.06	2.0	2.7	2.05	1.3	1.7	1.70	1.3
12 hr.	1.5	2.20	0.7	1.5	1.13	1.1	1.8	1.44	1.2
23 hr.	1.8	1.47	1.0	1.5	0.92	1.3	1.0	0.95	0.6

the end of the last shock impulse and the placing of *S* in the start box.

RESULTS

Table 1 presents median and mean latency for each of the nine groups on the test trial. Though latency raw scores are slightly skewed, the assumptions required for an application of the analysis of variance are not seriously violated. Consequently, this statistic was used. The results of this analysis are presented in Table 2. It is clearly apparent that the intensity of hunger is inversely related to latency ($F = 3.49$, $P = .05$). Emotionality failed to influence latency significantly, although there is a slight trend in the expected direction for the 0- and 23-hr. groups. There was also no significant interaction. A further test of the effects of emotionality, which is perhaps even more critical, compared the test performance of the 23-n, 23-w, and 23-s groups with their performance on the last training day. Groups 23-n and 23-w showed mean increases in latency on the test day of .68 sec. and .40 sec., respectively, whereas Group 23-s showed a mean decrease of 1.06 sec., but *t* tests indicated that these differences could be attributed to chance. A statistical comparison of the test scores of Groups 23-n, 23-w, and 23-s apart from the other groups is logically contraindicated by the results of the previous analysis of variance.

TABLE 2

ANALYSIS OF VARIANCE OF LATENCY SCORES

Source	*df*	*MS*	*F*
$\bar{D}$	2	17.75	3.49*
$\dot{D}$	2	7.50	1.47
$\bar{D} \times \dot{D}$	4	5.10	1.00
Within groups	117	5.09	
Total	125		

* $P = .05$.

Table 3 gives mean and median running time for each of the treatment groups on the test trial. The raw test scores were nonsymetrically distributed within cells and the between-cells variance was heterogeneous. In order to use the analysis of variance these data were converted to sten scores (**6**), which yielded normal distributions and homogeneity of variance between cells. Table 4 summarizes the analysis of variance. The *F* for $\bar{D}$ is significant beyond the .01 level; the *F* for $\dot{D}$ is less than one.

The interaction variance yields an *F* which is just short of the 2.45 needed for the .05 level. Reference to Table 3 shows this interaction derives mainly from the differential effects of emotionality upon the 23-hr. groups when compared to the other groups. It may be seen that running time is directly related to the intensity

TABLE 3

RUNNING TIME ON TEST TRIAL

$\bar{D}$ Food Deprivation	$\bar{D}$								
	No Shock (n)			Weak Shock (w)			Strong Shock (s)		
	Mean	*SD*	Mdn.	Mean	*SD*	Mdn.	Mean	*SD*	Mdn.
0 hr.	122.9	160.35	58.8	102.5	161.29	39.6	202.9	234.66	81.1
12 hr.	17.7	14.42	12.2	39.9	34.87	31.5	39.2	31.44	36.3
23 hr.	13.5	6.85	10.7	13.7	10.53	9.4	8.9	3.98	8.4

of emotionality for the 23-hr. *S*s, but tends to be inversely related for the 0- and 12-hr. *S*s. In this instance, the near significant interaction variance offers some justification for an independent analysis of the differences between the 23-hr. *S*s alone. Such an analysis appears even more appropriate if it is remembered that these *S*s were tested under conditions which most nearly simulated training. Therefore a simple analysis of variance was performed on the transformed data (sten scores) of groups 23-n, 23-w, and 23-s. The *F* of 2.34 with 2 and 39 *df* for between-groups falls far short of significance. The test performance of these groups was slightly poorer than that on the last training day. The median (raw score) increased from 6.4 to 9.9 sec. A simple sign test indicated this difference was statistically significant.

TABLE 4

ANALYSIS OF VARIANCE OF RUNNING TIME SCORES

Source	*df*	*MS*	*F*
$\bar{D}$	2	100.0	38.46**
D	2	0.5	
$\bar{D} \times D$	4	6.0	2.30
Within groups	117	2.6	
Total	125		

** $P = .01$.

DISCUSSION

Perhaps our most striking result is the effect $\bar{D}$ reduction had upon running time. The reduction of $\bar{D}$ to zero roughly doubles running time, and, significantly, four *S*s in these groups did not reach the goal box during the allotted 6 min. (A score of 6 min. was assigned to these *S*s for statistical purposes.) The effect produced by the reduction of $\bar{D}$ to 12 hr. is less pronounced. However, the mean running time for these groups combined is approximately equal to that at the beginning of training. Though there is a significant increase for the 23-hr. *S*s, this change is not appreciable in view of day-to-day variability. Whether the large decrement for the 0- and 12-hr. groups is due to a loss of the cue function or the energy function of $\bar{D}$ cannot be determined at this time.

To anticipate an argument which may be advanced, the change in procedure on the test day could have had a general disruptive effect upon performance. Such a general hypothesis could account for the over-all decrement in both running time and latency. Nevertheless, it would seem that some state of affairs must carry over to the maze situation in order to affect performance. Our only answer to this criticism would be that the test was specifically designed to minimize transfer of directive factors and to maximize energy transfer. Furthermore, a technique similar to the present one has been used successfully in several past studies (**3, 12, 15, 16**).

In general, the trend in the latency

scores on the test is similar to that of running time. However, the observed increase is not statistically significant. The behavior of Group 0-n proves interesting in view of Hull's original formulation of the generalized drive notion and the related Koch and Daniel study. It may be recalled that Hull extrapolated a curve depicting Skinner-box behavior to a point at which *S*s would be performing under zero relevant drive, and found that behavior strength was appreciably above zero. He attributed this residual behavior strength to the operation of $\dot{D}$. Later, Koch and Daniel (**11**) tested the validity of the extrapolation by placing *satiated* animals in a similar situation. They found that when $\dot{D}$ is reduced to zero, ${}_{s}\dot{E}_{R}$ is likewise reduced almost to zero. This finding is not clearly supported by the present data. Though the 0-n *S*s were very slow, 13 out of the 14 in this group reached the goal box within 6 min. Five of these *S*s were observed to nibble food. Qualitatively, however, the typical behavior of *S* in the maze following exposure to the shock apparatus did not appear to be goal directed. Specifically, they had a mincing gait in the alley, were inclined to urinate, and appeared acutely disturbed. For these reasons we tend to believe simple exposure to the shock apparatus induced mild emotionality, which, according to the generalized drive hypothesis, should have energized goal-directed behavior in this situation. It is felt that this particular result offers no clear evidence for or against the hypothesis.

The present data, viewed in their entirety, cannot be offered in support of Hull's generalized drive hypothesis. In those instances where $\dot{D}$ is low, $\dot{D}$ should facilitate behavior more than when $\dot{D}$ is high. The near-significant interaction obtained from the analysis of the running time data suggests a converse effect. It may be noted, by referring to Table 1, that median running time for the 23-hr. *S*s is directly related to the intensity of $\dot{D}$ whereas the relationship for the 0- and 12-hr. *S*s tends to be inverse. No appreciable interaction is evident in the latency data.

Several *ad hoc* explanations could be offered for the present negative findings. Perhaps the utilization of emotionality as a drive would bear careful scrutiny in view of the fact that many theorists have considered it a disruptive influence rather than a motivational state. Secondly, the possibility of competing responses cannot be ignored. Amsel (**2**) and Amsel and Cole (**4**) have clearly shown that, if competing responses are conditioned to cues associated with $\dot{D}$, the criterion behavior will undergo a decrement.

Finally, the criterion behavior may prove central to the problem. Those studies concerned with consummatory behavior have generally reported positive findings. On the other hand, positive evidence for a facilitative effect of $\dot{D}$ upon preparatory behavior is rare. Kendler (**10**) alone found positive evidence for a drive summation in a Skinner-box situation, and this finding was contingent upon the intensity of $\dot{D}$. Webb (**19**) reported positive results when the criterion behavior is door-opening behavior, although he used two need states which may be physiologically confounded. Amsel (**1**) failed to observe an increase in maze performance when a hunger need was added to a motivational complex in which the need state was emotionality. Behavior was facilitated when the dominant need state was changed to anxiety. Only Miller (**12**) reported positive results from a test which resembles the present one.

In conclusion, the present data cannot be offered in support of Hull's concept of generalized drive, i.e., substitution or summation. On the other hand, to propose that they constitute evidence against the hypothesis implies confidence in the fairness of the test, which we do not feel is entirely justified.

Summary

In order to test Hull's generalized drive hypothesis, 126 rats were trained to run a linear, alley maze while under a 23-hr. hunger need. Following 21 daily training trials, *S*s were assigned to nine treatment groups consisting of the possible combinations of 0, 12, and 23 hr.

food privation, as a relevant need, with zero, weak, and strong faradic stimulation as an irrelevant need. The "test" consisted of a trial similar to a training trial. The state induced by the faradic stimulation was called emotionality and viewed as possessing motivational properties. Both start-box latency and maze-running time were utilized as criterion behaviors.

The findings were mainly negative. The criterion behavior was not significantly influenced by the introduction of an irrelevant need into the motivational complex. As expected, changes in the relevant need state did significantly affect latency and running time. A near significant relevant need times irrelevant need interaction was observed. It was concluded that the present data provided no definite evidence either for or against the theory.

References

1. Amsel, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40,** 1–14.
2. Amsel, A. The effect upon level of consummatory response of the addition of anxiety to a motivational complex. *J. exp. Psychol.*, 1950, **40,** 709–715.
3. Amsel, A., & Maltzman, I. The effect upon generalized drive strength of emotionality as inferred from the level of consummatory response. *J. exp. Psychol.*, 1950, **40,** 563–569.
4. Amsel, A., & Cole, K. F. Generalization of fear-motivated interference with water intake. *J. exp. Psychol.*, 1953, **46,** 243–247.
5. Brandauer, C. M. A confirmation of Webb's data concerning the action of irrelevant drives. *J. exp. Psychol.*, 1953, **45,** 150–152.
6. Canfield, A. A. Sten scale. *Educ. psychol. Measmt*, 1951, **11,** 295–297.
7. Ellis, N. R., Siegel, P. S., & Wohl, S. A. The effect of emotionality upon the food intake of the rat. Paper read at Southeastern Psychol. Ass., Atlanta, May, 1955.
8. Finch, G. Hunger as a determinant of conditional and unconditional salivary response magnitude. *Amer. J. Physiol.*, 1938, **123,** 379–382.
9. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
10. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1954, **35,** 96–109.
11. Koch, S., & Daniel, W. H. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, **35,** 167–187.
12. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43,** 155–178.
13. Perin, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30,** 93–113.
14. Siegel, P. S. Alien drive, habit strength, and resistance to extinction. *J. comp. Psychol.*, 1946, **39,** 307–317.
15. Siegel, P. S., & Siegel, H. S. The effect of emotionality on the water intake of the rat. *J. comp. physiol. Psychol.*, 1949, **42,** 12–16.
16. Siegel, P. S., & Brantley, J. J. The relationship of emotionality to the consummatory response of eating. *J. exp. Psychol.*, 1951, **42,** 304–306.
17. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.
18. Strange, J. R. The effect of an irrelevant drive on the reaction tendency specific to another drive. *J. gen. Psychol.*, 1954, **51,** 31–40.
19. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39,** 1–14.
20. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23,** 506–522.
21. Wohl, S. A. The effect of emotionality upon the food intake of the rat. Unpublished master's thesis. Univer. of Alabama, 1955.
22. Zener, K. E., & McCurdy, H. G. Analysis of motivational factors in conditioned behavior: I. The different effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8,** 321–350.

(Received October 26, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

EMOTIONALITY AND THE YERKES–DODSON LAW

P. L. BROADHURST

Institute of Psychiatry (Maudsley Hospital), University of London

The Yerkes-Dodson Law (**17**) which states that the optimum motivation for a learning task decreases with increasing difficulty has been shown to hold for several species. See Young (**18**). It is the purpose of the present experiment to extend the range to include the rat. However, individual differences in drive strength are clearly important here, as may be seen from the current work showing the differential effect of anxiety on human learning (**16**). Recent evidence (**8**) supporting the identification of neuroticism as an autonomic drive (**7**) suggests that emotionality, as defined by defecation scores in Hall's open-field test (**10**), may be a relevant variable, since the lability of the autonomic nervous system probably underlies both neuroticism in humans and emotionality in rats (**6, 10**). Emotional rats will therefore be expected to show greater drive than nonemotionals in a situation in which the motivation used (air deprivation) is intense and of a kind likely to give rise to fear responses, and consequently to learn faster when the task is easy, but more slowly when it is hard. That is to say, the optimum drive level in the Yerkes-Dodson situation should be lower for the emotionals than for the nonemotionals. We may therefore predict an interaction between emotionality and the two variables of motivation and difficulty level which should themselves interact in the manner suggested by the Yerkes-Dodson Law.

METHOD

Apparatus.—The Y-shaped discrimination apparatus designed by Jonckheere (**12**) for surface swimming was adapted for underwater use. This is a conventional Y unit with metal walls 2 ft. high and alleys 4 in. wide. The stem of the Y is 7 in. long and the arms make an angle of 52° to the center line. Vertical partitions across each arm, 7 in. beyond the bifurcation, extend from top to bottom of the alley and contain the discrimination panels. The *S*s were forced to swim through the unit completely underwater by roofing in the alleys before the partitions with stainless steel hardware cloth just below the water level (9 in.). The discrimination panels were hung vertically and hinged below the surface of the water. They were made of frosted Perspex and each presented an illuminated area, 3½ in. square, at right angles to the line of sight from the bifurcation. Either could be locked into place, or left free, to be opened underwater by *S* to allow escape to the part of the alleys beyond the partitions which were not roofed in below the surface. From here, hardware cloth ramps gave access to a platform above water level.

The illumination was provided by two lamps in parallel in a 24-v. circuit. One lamp was located above water level behind each discrimination panel with its filament parallel to the panel. A fixed portion of a wire resistor could be switched into the circuit of either lamp in order to dim the intensity of the light shining through the adjacent panel, the other lamp being undimmed. The *S* was thus presented with a choice between a well lit and a less well lit avenue of escape from underwater. Three amounts of resistance were used; the greater the resistance the dimmer the light, yielding a greater difference between the illumination of the alleys, and an easier discrimination. Measurements with a light meter (Avo No. 2) indicated that this difference in illumination could be represented, in ascending order of difficulty, by the ratios 1:300, 1:60, and 1:15. These constitute the three levels of difficulty and are designated easy, moderate, and difficult.

A submersible cage, measuring 9 × 6 × 4½ in. wide, also made of stainless steel hardware cloth, was fitted with a sliding guillotine door which could be released from above the surface, thus allowing escape through the discrimination apparatus. Following a suggestion by Mason and Stone (**14**), the intensity of the air deprivation used as motivation was manipulated by detaining the *S*s for different lengths of time submerged

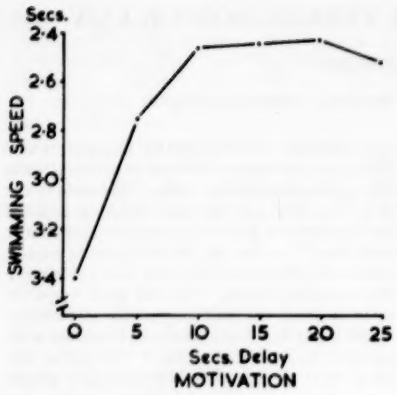

FIG. 1. The relationship between speed of swimming a 4-ft. straightaway underwater and intensity of imposed motivation (air deprivation) measured by the number of seconds delay underwater before release. Each point represents the mean time for 20 *Ss*; the data were collected on successive days.

underwater before releasing them to make the discrimination on which the promptness of their escape depended. Preliminary work, using 20 female rats as *Ss*, gave the results shown in Fig. 1. There is clearly no advantage in using delays longer than about 10 sec., since performance is not thereby improved. Indeed, delays longer than 20 sec. result in decreased swimming speed, presumably because of the effects of anoxia. Delays were therefore kept short, and delays of 0, 2, 4, and 8 sec. before release were selected, which represent approximately equal increments in swimming speed (see Fig. 1). These constitute the four levels of motivation.

Subjects.—Five replications of the $3 \times 2 \times 4$ factorial design (24 treatments) were used—that is, a total of 120 *Ss* was required. These 120 *Ss* were male albino rats which formed part of the second and third generations in a selective breeding study of emotionality being conducted in this laboratory (**3**). When *Ss* averaged 105.5 days of age ($SE \pm .25$), they were given the modified and standardized open-field test described in detail elsewhere (**1**). Briefly, *Ss* were exposed for 2 min. per day for four successive days in a circular arena 32¾ in. in diameter with white plywood walls 12½ in. high. A battery of loudspeakers and of photographic lamps above the arena provided sound ("white" noise) and light fields whose intensity at floor level in the arena averaged 78 db. (ref. .002 dynes/sq. cm.) and 165 cp., respectively. The average number of fecal boluses deposited per day constitute the emotional reactivity score, and only *Ss* scoring 3.3 or more were assigned to the emotional group (mean = 4.1, ± .15), and only *Ss* scoring 1.3 or less were assigned to the nonemotional group (mean = .8, ±.17). These two groups constitute the two levels of treatment in the emotionality variable; the difference between them is significant beyond the .1% level by *t* test.

The populations from which *Ss* were selected by virtue of their emotional elimination scores had been bred by brother × sister mating from a heterogeneous Wistar stock. They had been reared under standard conditions which are described fully elsewhere (**3**). In brief, these conditions featured controlled temperature and light/dark cycle in standard living quarters, standard conditions of husbandry—diet, routine care, number of animals per cage, etc.—and the minimum of handling. Some of them had been used for breeding purposes in the interval between testing in the open field and in the underwater discrimination unit, which interval itself varied with *Ss*' generation, but there is evidence (**2**) to suggest that neither the sexual experience, nor age differences of the order encountered, are likely to affect *Ss*' emotionality as measured. The mean age when experimentation began of the 45 *Ss* belonging to the second generation was 322.0 days (±6.09), and the mean age of the 75 *Ss* belonging to the third generation 213.4 days (±3.31). This difference is highly significant ($P < .001$ by *t* test); accordingly, the possible effect of this age difference upon the scores analyzed was investigated. There was no significant difference by *t* test between the over-all means of the two groups for the learning score used (discussed later), but, as might be expected, the older *Ss* swam significantly more slowly ($P < .01$, by *t* test). It need not, however, be anticipated that any systematic bias is thereby introduced into the analysis of the speed score, since *Ss* had been assigned to treatments randomly. As a check, a χ^2 test of the proportions of *Ss* of the two age groups assigned to each treatment combination showed no significant difference from that existing in the total group (45:75).

Procedure.—The random assignments of *Ss* to the 24 treatment combinations was made by forming the 60 emotional *Ss* into 12 groups of five each, equated as far as possible with respect to defecation scores, and the 60 nonemotional *Ss* into a further 12 groups of five each, similarly equated. The groups thus formed were then randomly assigned, within the emotionality dichotomy, to the various treatment combinations. The *Ss* were tested in groups of 24 (one complete replication) on 15 successive days.

The first five days were devoted to preliminary training during which *Ss* were successively given surface and underwater swimming practice in a 3-ft. straightaway, at which time the appropriate detention periods for the different *Ss* were established, then practice in the underwater discrimination unit to familiarize them with the operation of opening the doors formed by the discrimination panels. During this time the doors were unlit, and one side of the Y unit was blocked at the bifurcation by a solid partition, thus forcing the rat to one side or the other, in order to avoid the development of position habits. The side by which escape was thus permitted was varied randomly by use of a Gellerman series (**9**). At the end of this training all *Ss* were leaving the starting cage promptly on release, and opening the escape doors without difficulty.

On the next 10 days, 10 trials per day—a total of 100 trials—were given in the apparatus with a choice of alleys permitted and differences in illumination between the panels. No other illumination was present in the darkened room containing the water tank except for a small recording lamp and a radiant heat lamp above the self-draining metal boxes in which *Ss* were kept in the intertrial intervals. This interval was maintained relatively constant for all *Ss* despite the variation in delay ranging from 0 to 8 sec. by running them in larger or smaller groups, respectively. The brighter side of the unit was designated correct and kept open; an *S* choosing the darker side where the panel was locked was thereby forced to retrace its way to the brighter side in order to escape. That is, the correction method was used. The bright side was selected since it was constant for each of the three levels of difficulty, and because there was no tendency for *Ss* to prefer one side over the other. Thus, on the first trial upon which a choice was permitted, 56.7% went to the right-hand or bright side, and 43.3% to the dimmer one on the left. The difference from chance expectation yields a nonsignificant χ^2. The "correct" side was randomly varied from trial to trial by use of a selection of 10 Gellerman series, the same ones being used in the same order for all *Ss*. The water temperature was maintained at 20° C., the air temperature during testing averaged 19.5° C. (±.21).

The time from the moment when the door of the starting cage was released until *S*'s snout broke the surface of the water beyond the illuminated panel through which it had passed was recorded to the nearest $\frac{1}{5}$ sec. by stop watch, and errors, defined as any entry into the "incorrect" alley of *S*'s head and shoulders or more, noted. These constitute the time and error scores, respectively.

The five replications of the experiment were tested at intervals over 4 mo. and were followed by a partial replication which supplied substitutes for the three *Ss* which died during experimentation and the seven whose results were excluded because they developed position habits. A position habit was defined for this purpose as the choice on two days (20 trials) or more during the last five days of testing of one side—left or right—exclusively.

Results

The number of errorless trials out of the 100 trials given was counted for each *S* and the data subjected to a three-way analysis of variance after Pearson and Hartley's test (**15**) had disclosed no significant inhomogeneity of variance. The results are presented in Table 1. The *F* ratios for the difficulty and motivation main effects were calculated by using the Difficulty × Motivation interaction variance estimate, in view of the significance of the interaction between these two variables. The nature of this interaction is indicated in Fig. 2. Table 2 shows the mean scores and *SD*'s for the various levels of the two treatments having significant main effects. These data indicate the efficacy of the experimental manipulation of these two independent variables.

TABLE 1

Analyses of Variance of Learning and Speed Scores

Source	*df*	Number of Correct Trials		Average Time to Swim 21 In.	
		MS	*F*	*MS*†	*F*
Difficulty (D)	2	2260.2	23.5**	2.17	1.8
Motivation (M)	3	143.7	1.5	4.73	4.0**
Emotionality (E)	1	114.1	3.0	7.21	6.1*
D × M	6	96.1	2.5*	1.05	0.9
D × E	2	48.9	1.3	0.73	0.6
M × E	3	77.8	2.1	2.48	2.1
D × M × E	6	54.9	1.4	1.21	1.0
Ss	96	38.0		1.18	
Total	119				

* $P = .025$.
** $P = .01$.
† Units 1/10 sec.

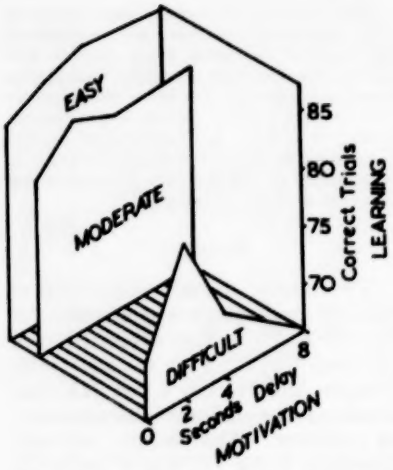

FIG. 2. A three-dimensional surface showing the relationship between learning scores in a discrimination task and (*a*) the intensity of the imposed motivation (air deprivation) measured by the number of seconds' delay underwater before release and (*b*) the level of difficulty of the task. The lamina are spaced to represent the over-all mean score for the appropriate difficulty level (see Table 2). Each point represents the mean score for 10 *Ss*.

Thus, each increase in the difficulty of the discrimination yields a significantly lower learning score, reaching the .01 level (by 1-tail *t* test) in the case of the easy vs. moderate difference, and the .001 level in the case of the moderate vs. difficult one. Increasing the time of air deprivation first increased the mean learning score significantly (0- vs. 2-sec. delay), then caused a significant decline from the peak value (2- vs. 8-sec. delay; $P < .05$ in both cases by *t* test). This latter finding is comparable with that of many workers reporting a curvilinear relationship between motivation and performance when increasingly intense motivation is employed (see also Fig. 1).

It is clear from these results that the Yerkes-Dodson Law may be taken as confirmed. The optimum motivation for a discrimination task demonstrably decreases with increasing difficulty of the task. This effect, represented by the first-order interaction between the difficulty and motivational treatments, is significant at a satisfactory level, and part of the prediction is thus verified. It is equally clear that the prediction, as made above, relating to the effects of emotional reactivity on this complex relationship is not fulfilled. The trend of the results is partly in the predicted direction. Thus emotionals learn the easy discrimination faster than nonemotionals, and are about equal on the moderately difficult one, and similarly learn faster under low

TABLE 2

BREAKDOWN OF LEARNING SCORES

Motivation Level (Air Deprivation)	Difficulty of Discrimination							
	Easy		Moderate		Difficult		All	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0 sec.	84.8	6.6	81.3	6.3	71.1	9.0	79.1	9.4
2 sec.	86.4	4.7	84.7	5.9	79.5	4.8	83.5	6.0
4 sec.	87.7	3.9	83.0	7.1	71.6	6.2	80.8	9.0
8 sec.	86.8	3.9	83.2	5.9	66.1	6.5	78.7	10.6
All	86.4	5.0	83.1	6.5	72.1	8.3	80.5	9.1

motivation, and are about equal at medium levels. But on the difficult discrimination or under the most intense motivation used they again show a slight superiority. Consequently, the emotional *Ss* show an over-all superiority in learning, which, however, does not reach significance, and this is reflected in the failure of the one second-order and the two remaining first-order interactions, all of which involve the emotionality variable, to reach significance.

If it is conceded that the curvilinear relation involved makes such a procedure permissible, the partial success of the prediction relating to drive level and emotionality can readily be demonstrated. An analysis of variance based on the first two lowest levels of motivation only—that is, before the decline in learning associated with motivation greater than the optimum begins (see Fig. 2)—shows two significant effects. The first is the expected and highly significant one associated with level of difficulty, the second is the first-order interaction between Motivation and Emotionality ($P < .05$). This effect may be summarized by saying that the emotional *Ss* show superior learning under the minimal motivation used, but that this advantage disappears with the increase of motivation from 0 to 2 sec. delay before release.

Another way to investigate the relation of emotionality to drive level is to consider the *speed* of the response made to the motivational stimulus, irrespective of whether or not it resulted in improved learning. In order to derive a speed score for each *S*, and one which reflects its speed of swimming independent of the degree of learning attained, the time scores for the first day upon which all 10 trials were recorded as correct were averaged. In this way the individual *Ss*' speeds may be compared at exactly comparable points on their respective learning curves. Since the distance swum in the apparatus is essentially the same for each *S* making a "correct" choice (21 in.), the time scores can be used directly for the purpose, as was done in preparing Fig. 1, thus avoiding the difficulty relating to averaging time over distance scores discussed by Crespi (**5**). For the 13 *Ss* assigned to the difficult discrimination who never learned it to perfection, the average of the last day's trials was calculated, omitting times for those trials when incorrect choices were made. An analysis of variance performed on these data after no significant inhomogeneity of variance had been demonstrated gave results which are also shown in Table 1. From this analysis it will be seen that the superiority of the emotional *Ss* over the nonemotional *Ss* in speed of swimming reaches satisfactory significance. The mean time scores are 1.80 and 1.96 sec., respectively. Inserting these values on a plot of swimming speed against motivational delays (like Fig. 1) shows that the difference of .16 sec. is equivalent to the effect on over-all speed scores of an increase in motivational delay of about 4 sec. Moreover, this effect is not significantly associated with any of the other variables in the analysis, from which it can be concluded that the emotionally reactive rats display a characteristically higher level of drive, which, within the limits of the present experiment, is not affected by the intensity of the motivation imposed. It may also be noted from this analysis that the absence of a significant effect for Difficulty shows that the speed score selected was in fact independent of the degree of learning, and that the significance of Motivation confirms the

efficacy of the experimental manipulation of the air deprivation variable. Thus, the mean time taken to swim 21 in. underwater decreased progressively from 1.98 sec. for 0 sec. delay to 1.97 for 2 sec., to 1.85 (4 sec.), and finally to 1.71 (8 sec.). The *SD*'s are .48, .33, .31, and .25 sec. respectively. Only the differences between each of the first three levels (0, 2, and 4 sec.) and the last one (8 sec.) reach significance (.05 level or beyond by 1-tail *t* test).

Discussion

The confirmation of the Yerkes-Dodson Law reported here gains in interest because it employed a different species from those hitherto used in demonstrating the principle, and because it occurred under rather different experimental conditions. The motivation in particular was very different from that used in any of the other studies of the Yerkes-Dodson Law; the electric shock previously employed as an aversive stimulus was eschewed because of the difficulty of controlling it, and because of its disruptive effect, especially when strong, upon early learning. Air deprivation as used shares with shock the quality of being a rather intense stimulus, and the advantage of being manipulable within the testing situation. It has, in addition, been shown in practice to be easily varied in intensity.

It is true that in both this and the original study a visual brightness discrimination was used as the method of varying the degree of difficulty of the learning tasks. The resemblance hardly goes beyond the semantic, however. The similarity between the actual pieces of apparatus used is slight. The Yerkes box has been described as "spectacularly inefficient" (**13**), whereas the underwater discrimination technique as used in this study gave rapid learning without undue preliminary training.

The results obtained from the original study and the present one are surprisingly similar. A comparison of Fig. 2 with that given by Yerkes and Dodson (**17**, p. 479) shows that the optimum motivation is in general slightly lower in the scale used than that encountered by them, and that there is no level of difficulty for which an optimum was never reached as was the case with their easiest discrimination. An increasing sharpness of the optimum peak with increasing difficulty is characteristic of both graphs. The dissimilarities mentioned can all be ascribed to differences in the levels of motivation selected. The relation of motivation to degree of learning is typically curvilinear, a decrease in learning following motivation more intense than the optimum, so it is not always easy to select suitable levels of motivation to reveal the optimum learning, especially if it is required to include different optima associated with different levels of difficulty, and more especially with a novel motivation like air deprivation about which little is so far known.

The confirmation of the Law is the more striking because of this difference in motivation. The needs arising from electric shock and from air deprivation are different physiologically, though they both give rise to drives mediated, in part, by fear. The further generalization of the Law must await experimentation using other situations and other drives, but the outlook is promising in view of the analogies in the human field already existing.

The possible role of anoxia must not be overlooked in connection with the use of air deprivation, since it is known (**11**) that it can cause learning defects comparable to those caused by electroconvulsive shock. It seems unlikely, however, that air deprivation of the duration used here can have led to anoxic damage to the nervous system sufficient to account for the curvilinear relation with learning. To the arguments which lead Mason and Stone (**14**) to reject this explanation of their comparable findings, we may add the following germane considerations from this experiment. Firstly, if anoxia were important in this connection, it would be reasonable to anticipate that all the groups subjected to

the greatest air deprivation would show a decrement in learning. A glance at Fig. 2 shows that this is not the case. Secondly, the analysis of the speed of swimming showed no significant difference between the difficulty levels (see Table 1). Now, the *S*s assigned to the difficult discrimination had, at the time when the speed scores were selected, experienced much more air deprivation than had the other groups. They had made many more errors obliging them to hold their breath longer before escape, and they had also undergone many more trials. Neither of these circumstances, likely to increase any deficit ascribable to anoxia, depressed swimming speed, however.

The drive characteristics of emotionality as defined by open-field defecation scores are only partly elucidated by this experiment. Emotionally reactive *S*s have a significantly higher speed of swimming—that is, they respond more vigorously to the same degree of imposed motivation. To this extent, they may be regarded either as having a higher drive level in general or as more susceptible to drive arousal. Further work to investigate the speed of swimming of these *S*s, and employing a situation designed to demonstrate the Crespi effect (**5**) consequent upon changes in drive level, confirms the superior speed of the emotionals and the presence of the "elation" effect but indicates that they do not respond more to a sudden increase in drive than do the nonemotionals (**4**). It thus seems that drive level rather than drive arousal is involved.

Yet this higher drive level did not in general effect learning in this experiment in the manner expected. It seems unfruitful to enter into detailed speculations about the reasons for this seeming paradox—suffice it to say that it is probable that the situation used here is too complex to expect that any definitive solution may be found in the present data. Nevertheless the general finding relating to drive level does have some bearing on the postulated relation between emotionality in rats and neuroticism in humans. This cross-species identification is not directly verifiable, but it can be cautiously said that the present results, at the very least, do nothing to render it improbable.

Summary

The experiment was designed to test the validity of the Yerkes-Dodson Law and to investigate some of the drive characteristics of emotionality of rats. In a $3 \times 4 \times 2$ factorial design having five replications, 120 male albino rats were used as *S*s. Three levels of difficulty of an underwater brightness discrimination, four levels of motivation deriving from different degrees of air deprivation, and two levels of emotionality defined in terms of defecation scores on the open-field test were used. Time and error scores from 100 trials were secured.

Analysis of variance of the results shows that the Yerkes-Dodson Law, as demonstrated by an appropriate interaction between difficulty and motivation, is confirmed at an acceptable level of significance. The prediction relating to the effects of emotionality on motivation is only fulfilled in part; a significantly higher drive level of emotionals as shown by their swimming speed is shown.

References

1. Broadhurst, P. L. Determinants of emotionality in the rat: I. Situational factors. *Brit. J. Psychol.*, 1957, **48**, 1–12.
2. Broadhurst, P. L. Determinants of emotionality in the rat: II. Antecedent factors. *Brit. J. Psychol.*, in press.
3. Broadhurst, P. L. Emotionality in the rat: a study of its determinants, inheritance and relation to some aspects of motivation. Unpublished doctor's dissertation, Univer. London, 1957.
4. Broadhurst, P. L. A note on a "Crespi effect" in the analysis of emotionality as a drive in rats. *Brit. J. Psychol.*, in press.
5. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
6. Eysenck, H. J. *The structure of human personality.* London: Methuen, 1953.
7. Eysenck, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.
8. Eysenck, H. J. Reminiscence, drive and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.
9. Gellerman, L. W. Chance orders of alternating stimuli in visual discrimina-

tion experiments. *J. genet. Psychol.*, 1933, **42**, 206–208.

10. HALL, C. S. Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. comp. Psychol.*, 1934, **18**, 385–403.
11. HAYES, K. J. Anoxic and convulsive amnesia in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 216–217.
12. JONCKHEERE, A. R. A study of "fixation" behaviour in the rat. Unpublished doctor's dissertation, Univer. London, 1956.
13. MCCLEARN, G. E., & HARLOW, H. F. The effect of spatial contiguity on discrimination learning by Rhesus monkeys. *J. comp. physiol. Psychol.*, 1954, **47**, 391–394.
14. MASON, W. A., & STONE, C. P. Maze performance of rats under conditions of surface and underwater swimming. *J. comp. physiol. Psychol.*, 1953, **46**, 159–165.
15. PEARSON, E. S., & HARTLEY, H. O. *Biometrika tables for statisticians*. Vol. 1. Cambridge, Eng.: Cambridge Univer. Press, 1956.
16. TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.
17. YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *J. comp. Neurol. Psychol.*, 1908, **18**, 459–482.
18. YOUNG, P. T. *Motivation of behavior*. New York: Wiley, 1936.

(Received October 30, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

TRANSFER FROM VERBAL PRETRAINING TO MOTOR PERFORMANCE AS A FUNCTION OF RESPONSE SIMILARITY AND ANGLE OF MOVEMENT[1]

DONALD R. HOFFELD

University of Wisconsin

The effects of verbal pretraining upon motor performance have been the subject of considerable research since 1950. In general the procedure in these experiments has been to give one or more groups some type of verbal pretraining, usually consisting of verbal responses to a physical stimulus, and then to compare the performance of these pretrained groups to a control group which has had no pretraining. The results of studies in this area have been conflicting to date. Using simple discrimination tasks, several studies (**1, 9, 13, 14**) have found evidence for significant positive transfer effects, while several studies using more complex tasks (**12, 16**) have found no positive transfer effects. This situation was clarified to some extent by Battig (**3**) who found that positive transfer decreased as task complexity increased. A number of variables influencing the amount of transfer from verbal pretraining still remain unclarified, however.

The purpose of this experiment was to determine the effect of difficulty of a motor task and the relevancy of responses made during pretraining upon the amount of transfer. The mirror tracing task, first used by Starch (**15**), was selected because of the relatively simple response and the relative simplicity of verbalizing the motor response. Preliminary work indicated that angle of tracing might be a difficulty dimension, so the angle between the mirror and the tracing pattern was selected as an additional variable. The responses made during pretraining consisted of drawing lines of 0° and 90° in response to verbal stimuli. It is assumed that all *S*s were at a plateau of performance for the 0° angle before entering the experiment and that consequently the practice on the 90° angle is the only important response practice gained from the pretraining. In view of this, the relevancy of the response during pretraining varies directly with the angle, i.e., the larger the angle, the more relevant the response. The selection of these variables then permits the analysis of the relation between task performance and angle of tracing, effect of relevancy of pretraining upon transfer, and the overall transfer effect.

METHOD

Apparatus.—The mirror tracing apparatus consisted of a rectangular box 15¼ in. long, 10 in. high, and 15 in. deep. A window was cut in the rear of the box and a one-way mirror fixed in place so that *E* could observe the progress of *S* at all times. The interior of the box was illuminated by two 8-w. florescent tubes. Shields were located so as to prevent the light from shining directly into *S*'s eyes and to prevent *S* from seeing the tracing pattern without aid of the mirror. Tracing patterns were inserted into the box on masonite sheets by *E*. The tracing patterns consisted of a series of patterns com-

[1] This research was supported in part by a grant made by the National Science Foundation to Dr. W. J. Brogden, and is based on a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at the University of Wisconsin. The author wishes to express his appreciation to Dr. W. J. Brogden, under whose guidance this research was done, for his helpful suggestions.

posed of 0° and 15°, 30°, 45°, 60°, or 75° angles. The tracing pathway consisted of a 0° section ½ in. in length followed by a ½-in. section at an angle of −15°, −30°, −45°, −60°, or −75°, followed by another ½-in. section of 15°, 30°, 45°, 60°, or 75°. This sequence of sections was repeated eight times and the pathway concluded by a final ½-in. 0° section. The pathway was approximately ¼ in. wide. The patterns were reproduced on 8½ × 13-in. sheets of paper by the Xerox process. The mirror tracing box was placed directly in front of *S*. The posture of *S* was fixed by use of an adjustable chin rest. A mechanical pencil was used for tracing the pattern. The arrangement was such that *E* sat directly across from *S* and observed the course of a trial through the rear of the box.

Subjects.—The *S*s were 150 undergraduate psychology students who received class points for their services. They are assigned to groups on the basis of a pretest consisting of a random pattern containing one each of the five angles used in the experiment. The pattern was constructed in the same manner as described above, but contained only one sample of each of the five angles. In all, 159 *S*s were pretested and 150 selected for the experiment on the basis of time scores. Approximately the first 125 *S*s were assigned randomly to groups until one group had been completed and the remainder of the assignments made in a manner to match the groups' pretest time scores for means and *SD*'s. Since this matching procedure was used, no attempt was made to equate the groups for sex. All *S*s were right-handed.

Procedure.—The experimental paradigm used is as follows: Phase 1, both experimental and control groups took the pretest; Phase 2, the experimental group received relevant verbal pretraining while the control group received nonrelevant verbal pretraining; Phase 3, both experimental and control groups received 10 test trials.

Instructions for Phase 1 were given telling *S* that the task was to trace through the pattern as quickly as possible, but cautioning him to remain within the pathway at all times, and to return to the pathway if his pencil left it. After completion of Phase 1, *S*s were assigned to groups as described above.

For Phase 2, all *S*s in the control group were given nonrelevant verbal pretraining consisting of learning a list of six nonsense syllables. The *S*s in the experimental group learned word association pairs verbally. The three association pairs used were away–toward, right–right, and toward–away. These were typed on 3 × 5-in. cards and held at eye level to *S*. The words were learned to a criterion of one perfect trial. The *S* was then seated so that he had directly in front of him a blank sheet of paper, and, in response to the first word of each association pair, drew lines to symbolize the second word. A sequence of stimulus words was given until *S* had made eight consecutive correct responses. This completed training for the experimental group. After learning the nonsense syllables, the control group was told that the list had nothing to do with the rest of the experiment and was merely a rest interval activity. Both experimental and control groups then received a review of the original instructions, with both groups being reminded of the mirror reversal effect.

During Phase 3, the experimental (E) and control groups (C) were divided into five subgroups each, with each of the five angles being traced by an experimental and a control subgroup. The five angles used are given above, and the angle is measured between the frontal plane of the mirror and the orientation of the tracing pattern lines.

Each *S* then traced 10 samples of the same pattern. Time scores were recorded on a stop watch operated by *E*. At the conclusion of each trial, *E* removed the masonite insert and substituted another sample of the same pattern. The interval between trials was approximately 10 sec. During the course of the test trials, all *S*s were told on Trial 5 that they were improving. No other knowledge of results was given.

Time scores were used as the dependent variable in part because *S*s were matched on this variable. Error scores were not used because of the difficulty of objective scoring and the phenomenon of blocking in mirror tracing (see Hoffeld (**11**).

Results

Bartlett's test for homogeneity of variance (**7**) was applied to the time scores. The resulting chi-square value is 490.64 ($P < .001$), which indicates significant heterogeneity of variances. An inspection of the data indicated that a logarithmic transformation could be used to reduce heterogeneity and help normalize the distribution. All time scores were transformed into log seconds. The resulting chi-square value is 127.99 ($.05 > P > .025$). An inspection of the variances revealed that this significant chi-square value was due to one variance (E-75° on Trial 8). If

TABLE 1

ANALYSIS OF VARIANCE FOR TIME SCORES IN LOG SECONDS

Source	*df*	*MS*	*F*
Trials (T)	9	1.28816	44.65**
Linear	1	10.09985	442.01**
Quadratic	1	1.01151	35.06**
Cubic	1	.31111	10.78**
Quartic	1	.04064	1.41
Quintic	1	.02868	.99
Residual	4	.02541	.88
Pretraining (P)	1	.00237	.08
Angle (A)	4	1.00285	34.76**
A × P	4	.07397	2.56*
Experimental	4		
Linear	1	2.47693	85.86**
Quadratic	1	.44682	15.49**
Cubic	1	.01882	.65
Quartic	1	.03841	1.34
Control	4		
Linear	1	.73235	25.38**
Quadratic	1	.44699	15.49**
Cubic	1	.00510	.18
Quartic	1	.14181	4.92*
A × T	36	.01018	.35
T × P	9	.00668	.23
A × T × P	36	.00746	.26
Error	1400	.02885	
Total	1499		

* $P = .05$.
** $P = .01$.

this variance were not included in the test, the chi-square value would be 108.08 ($P > .05$). In view of this fact and also because of the sensitivity of the Bartlett's test, the .05 level of significance was used in all *F* tests.

An analysis of variance was performed on the transformed time scores and is summarized in Table 1. The significant *F* for trials indicates a general improvement in the performance of *S*s. The absence of any significant interactions involving trials indicates that the form of the learning curves is the same for all groups. The data for all experimental and control groups were therefore combined into a single group curve which was fitted by orthogonal polynomial analysis (**8**). The breakdown of the orthogonal polynomial analysis is also given in Table 1. The over-all group curve shows significant linear, quadratic, and cubic components. The significant *F* for the angle main effect indicates a difference in performance between the different angles. The angle main effect and the angle by pretraining interaction have been broken down by orthogonal polynomial analysis, with separate curves having been fitted to the experimental and control groups (**10**). The empirical curves are shown in Fig. 1. The experimental group has significant linear and quadratic components, while the control group has significant linear, quadratic, and quartic components. Tests of the differences between means by the Duncan Range Test (**6**) indicate that the experimental group for 75° (E-75°) is significantly ($P = .05$) better in performance than E-30°, C-60°, and E-15°, and that C-75° is significantly better in performance than E-30°. No other differences in means are significant. However, in view of the fact that the variable of angle is evenly and quantitatively spaced, the orthogonal polynomial is the more powerful test, and the breakdown indicates that differences between means in the control and experimental groups exist but do not show up in the Duncan Range Tests.

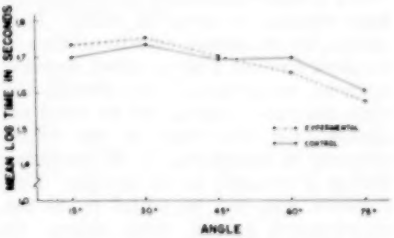

FIG. 1. Curves for the experimental and control groups for time scores in log seconds as a function of angle of movement.

Discussion

Performance as a function of degree of relevancy of relevant S-R pretraining is shown to be affected by the presence of a significant Pretraining × Angle interaction. For the angles (15°, 30°, and 45°) for which the response made during pretraining was less relevant, there appears to be an inhibitory effect. This is consistent with the results of Battig (2) and Battig, Hoffeld, Seidenstein, and Brogden (3), both of whom found nonsignificant inhibitory effects of relevant-*S* pretraining. For the angles (60° and 75°) where the pretraining was more relevant, there seems to be a facilitory effect.

The orthogonal polynomial analysis indicates that the differences between angles for the control group are significant, and inspection of the curve in Fig. 1 shows that there was a general tendency for performance to be higher for the easier angles for the groups having relevant S-R pretraining. These results are also consistent with those of Battig (2), who found that positive transfer decreased with increased task difficulty.

The reason for the inhibitory effect of the verbal pretraining in some cases and the facilitory effect in other cases may be explained in terms of relevance of pretraining and resulting interfering tendencies. It was found in a pilot study that it was not possible to have *S*s verbalize overtly at each choice point, since the time to make the verbal response tended to slow down performance too much. This is consistent with the results of Rossman and Goss (14) who found a slight inhibitory effect when *S*s were required to verbalize aloud. It is reasonable to assume that some time is required for *S* to make a verbal response covertly, even though this time is less than the time required for an overt response. Let this time for the verbal response be designated by K, assuming that it is constant for all groups. This response time will be added to all the experimental groups' time scores. In addition, there is a quantity y, which is the decrease in time score as a result of the verbal response having been made and serving as a stimulus for the proper motor response. This quantity y is then always negative with respect to the time scores. In addition, there is a quantity z, which is a function of the particular angle used in the tracing pattern, and a quantity e, which is a random error component. Thus $\textit{Time Score} = z - y + K + e$. It is apparent that the relevancy of the response's decrease will effect a decrease in y. Thus when y is less than K, an inhibitory effect appears. The data are consistent with this explanation.

The data on performance as a function of angle of movement are roughly in accord with those of Corrigan and Brogden (5), who found a symmetrical sine function around 45°, with the poorest performance coming for the 45° angle. The task used by Corrigan and Brogden was a constant velocity tracking task with the angle of tracking as the independent variable.

Using the same task as Corrigan and Brogden, Brogden (4) found that performance varied over trials for the different angles used, with no learning being evidenced for some angles. Absence of an angle by trials interaction in this present experiment indicates that this does not hold true for the mirror tracing task. Such differences are probably due to the fact that in this research *S*s repeatedly had to change direction of movement, while *S*s in Brogden's study merely tracked a pip moving at a constant velocity on a straight track.

Summary

The purpose of this research was to determine whether the amount of transfer from verbal pretraining to a mirror tracing task varies differentially as a function of angle of movement. All groups were matched in time scores on the basis of a mirror tracing pretest. The dependent variable was time to trace a pattern made up of horizontal lines and lines forming angles with the horizontal lines. The experimental group had verbal pretraining of a relevant S-R nature, consisting of learning to associate 90° movements with spoken stimulus-response pairs. The control group learned nonsense syllables as nonrelevant verbal pretraining. Each of the

groups was then divided up into five subgroups, with each subgroup tracing patterns made up of one of five angles.

A logarithmic transformation was performed on the time scores because of heterogeneity of variance. The results of the subsequent analysis show a differential effect of pretraining for the different angles, with a slight inhibitory effect appearing for angles 15°, 30°, and 45° and a facilitory effect appearing for angles 60° and 75°. An analysis by orthogonal polynomials indicated significant linear and quadratic trends for the experimental group data. Performance is also found to be a function of angle of movement for the control group. A curve fitted by orthogonal polynomials was found to have significant linear, quadratic, and quartic components.

It is concluded that transfer from verbal pretraining to a mirror tracing task varies directly as a function of similarity of responses made during pretraining. It is also concluded that performance varies as a function of angle of movement.

References

1. Baker, K. E., & Wylie, R. C. Transfer of verbal pretraining to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
2. Battig, W. F. Transfer from verbal pretraining to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51**, 371–376.
3. Battig, W. F., Hoffeld, D. R., Seidenstein, S., & Brogden, W. J. Supplementary report: effect of verbal pretraining on the acquisition of a complex motor skill. *J. exp. Psychol.*, 1957, **54**, 375–376.
4. Brogden, W. J. The trigonometric relationship of precision and angle of linear pursuit movements as a function of amount of practice. *Amer. J. Psychol.*, 1953, **66**, 45–56.
5. Corrigan, R. E., & Brogden, W. J. The trigonometric relationship of precision and angle of linear pursuit movements. *Amer. J. Psychol.*, 1949, **62**, 90–98.
6. Duncan, R. B. A significance test for differences between rank treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2** (n. s.), 171–189.
7. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
8. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* (4th Ed.) New York: Hafner, 1953.
9. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
10. Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.
11. Hoffeld, D. R. The effect of angle of movement and similarity of responses made during verbal pretraining upon performance in a mirror tracing task. Unpublished master's thesis, Univer. of Wisconsin, 1956.
12. Lawshe, C. H., & Cary, W. Verbalization and learning a manipulative task. *J. appl. Psychol.*, 1952, **36**, 44–46.
13. McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.
14. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
15. Starch, D. A demonstration of the trial and error method of learning. *Psychol. Bull.*, 1910, **7**, 20–23.
16. Spieth, W. A. A study of the effects of verbal pretraining on the subsequent learning of two mutually antagonistic motor tasks in alternating practice. Unpublished master's thesis, State Univer. of Iowa, 1951.

(Received November 1, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

THE DISCRIMINATION OF SPEECH SOUNDS WITHIN AND ACROSS PHONEME BOUNDARIES[1]

ALVIN M. LIBERMAN, KATHERINE SAFFORD HARRIS, HOWARD S. HOFFMAN, AND BELVER C. GRIFFITH[2]

Haskins Laboratories and the University of Connecticut

In listening to speech, one typically reduces the number and variety of the many sounds with which he is bombarded by casting them into one or another of the phoneme[3] categories that his language allows. Thus, a listener will identify as *b*, for example, quite a large number of acoustically different sounds. Although these differences are likely to be many and various, some of them will occur along an acoustic continuum that contains cues for a different phoneme, such as *d*. This is important for the present study because it provides a basis for the question to be examined here: whether or not, with similar acoustic differences, a listener can better discriminate between sounds that lie on opposite sides of a phoneme boundary than he can between sounds that fall within the same phoneme category.

There are grounds for expecting an affirmative answer to this question. The most obvious, perhaps, are to be found in the common experience that in learning a new language one often has difficulty in making all the appropriate sound discriminations. The evidence for this is impressionistic in the extreme, and there is little information that would permit a definition of the more specific aspects of the difficulty. In whatever degree this difficulty exists, however, a reasonable assumption is that some part of it arises from the fact that a person who is newly exposed to the sounds of a strange language finds it necessary to categorize familiar acoustic continua in unfamiliar ways. If his discriminations have, by previous training, been sharpened and dulled according to the position of the phoneme boundaries of his native language, and if the acoustic continua of the old language are categorized differently by the new one, then the learner might be expected to have difficulty perceiving the sounds of the new language until he had mastered some new discriminations and, perhaps, unlearned some old ones.

In more explicit psychological terms, an affirmative answer is to be expected on the basis that the situations being considered here clearly meet the conditions for acquired similarity and acquired distinctiveness. If either or both of these processes do, in fact, occur, then two speech sounds which a listener normally lumps into the same phoneme class would come to be less discriminable than sounds to which he habitually attaches different phoneme labels. Indeed, one might conceivably find in language some very common and

[1] This research was supported in part by the Carnegie Corporation of New York and in part by the Department of Defense in connection with Contract DA49-170-sc-1642.

[2] Now at Bell Telephone Laboratories, Murray Hill, N. J.

[3] The phoneme is most often taken to be the smallest unit of speech that can, by itself, distinguish one utterance from another as to meaning. Thus, the existence of the two words *bill* and *dill* makes it clear that *b* and *d* are different phonemes in English. It should be emphasized that the phoneme is not a single sound, but is, rather, a class which can and usually does include a great many sounds that differ from each other in various ways without causing any change in meaning.

easily accessible cases in which the effects of such processes as acquired similarity and acquired distinctiveness are as great as many years of practice can make them.

The present experiment was designed to investigate the relation between phonemic labeling and discrimination in one language and within one group of phonemes. For this purpose a synthesizer was used to generate speech-like sounds and to vary them in small steps along an acoustic continuum known to contain important cues for the perception of the voiced stops, *b*, *d*, and *g*. When listeners are asked to label these sounds as *b*, *d*, or *g*, they normally tend by their responses to divide the continuum into three sharply defined phoneme categories, the shifts from one response, or phoneme label, to another being very abrupt. It was the purpose of this experiment to determine how well these same sounds can be discriminated, and, in particular, to see whether the discrimination functions have sharp inflections that correspond in position to the abrupt shifts (i.e., phoneme boundaries) in the labeling responses. In addition, an attempt has been made to determine to what extent the relation between discrimination and labeling has here been reduced to its theoretical limit. For that purpose, the obtained discriminations have been compared with a function that is computed from the labeling data on the extreme assumption that the listener cannot hear any differences among these sounds beyond those that are revealed by his use of the phonemic labels.

Method

Apparatus.—A special-purpose instrument, called a pattern playback, was used to generate the stimuli of this experiment. This instrument, which has been described in earlier papers (**1, 2, 3**), converts hand-painted spectrograms into sound, thus making it possible to synthesize speech-like auditory patterns and to control the various aspects of the pattern quite precisely.

Stimuli.—Figure 1 illustrates the spectrograms used to produce the stimuli. The stimulus variable is the direction and extent of the second-formant transition, this variable having been found previously (**6**) to be important for the perceived distinctions among *b*, *d*, and *g*. In the stimulus pattern at the extreme left of the top row of the figure, the second formant rises from a point 840 cps below its steady-state level, and in the pattern at the extreme right of the bottom row it falls from a point 720 cps above the steady state. Between these two extremes, the starting point of the transition varies in steps of 120 cps. For convenience these stimuli will be referred to by number, from 1 through 14, as indicated in Fig. 1.

The rising transition of the first formant had been found previously to be a marker for the class of voiced stops (**4**) and, as can be seen in Fig. 1, this first-formant transition is constant in all the stimuli. In the steady-state part of the pattern the first formant centers at 360 cps and the second at 2160 cps. Formants at these frequencies produce a synthetic approximation to the vowel *e* (as in *gate*).

The spectrograms were converted into sound on the pattern playback and recorded on magnetic tape. By copying, cutting, and splicing the magnetic tape, all the stimulus arrangements described below were made.

Stimulus presentation and Ss.—The stimuli were presented to *Ss* in two ways; singly, to determine how *Ss* would label them as *b*, *d*, or *g*, and in an ABX arrangement to determine to what extent *Ss* could discriminate them on any basis at all.

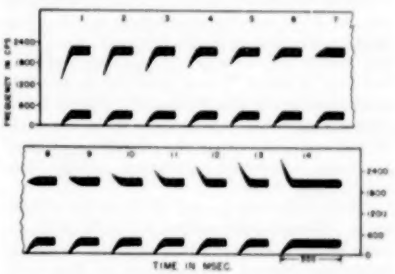

Fig. 1. Illustrations of the spectrographic patterns from which the stimuli of the experiment were produced. Pattern 14, at the lower right, is complete in all respects.

For the labeling part of the experiment, six magnetic tapes were prepared, each of which contained a different randomization of the entire series of 14 stimuli. There was a 6-sec. interval between stimuli. These tapes were presented to *Ss* with instructions to judge each stimulus as *b*, *d*, or *g*.

To test discrimination by the ABX procedure, the stimuli were arranged in triads, each of which consisted of an A stimulus, a B stimulus, and a third stimulus, X, which was identical either to A or to B. The *Ss* were instructed to determine whether X was the same as A or the same as B. It was strongly emphasized to each *S* that he was to make his judgment on the basis of any cues he could hear. The measure of discriminability is, of course, the percentage of the time that the X stimulus is correctly matched to A or to B.

It was not the purpose of this experiment to obtain actual DL's, but only to measure relative discriminability at every step on the continuum. For that purpose, the A and B stimuli of the ABX triad were made up by pairing each stimulus with the stimulus one step, two steps, and three steps removed from it. Each stimulus was paired in this fashion with the stimuli lying to its right on the continuum shown in Fig. 1. Thus, Stimulus 1 was paired with Stimulus 2 to form the A and B stimuli, respectively, of one ABX triad. Stimulus 1 was paired with Stimulus 3 in another triad, and with Stimulus 4 in a third. Similarly, Stimulus 2 was paired with Stimulus 3, with Stimulus 4, and with Stimulus 5, and so on for the remaining stimuli on the continuum. There was no stimulus one step to the right of Stimulus 14, so 14 does not appear as an A stimulus in the one-step series. For analogous reasons, Stimuli 13 and 14 do not appear as A stimuli in the two-step series, and Stimuli 12, 13, and 14 are missing from the three-step series.

From the above discussion, it follows that the number of A and B combinations was 36. For each combination of A and B stimuli, there were two triads—one in which X was identical to A and one in which X was identical to B. The total number of triads was 72. These triads were arranged into six tapes of 12 triads each. Within each ABX triad, the stimuli were spaced at 1-sec. intervals; successive triads were separated by 10 sec.

There were two groups of *Ss* in the experiment. Since the procedures for the two groups were slightly different they will be described separately. Group I consisted of five paid volunteers, all undergraduate students at the University of Connecticut, who had never heard any synthetic speech prior to this experiment. For the first 17 sessions, *Ss* were given only the ABX discrimination task to do, without being told that these were synthetic speech sounds. Beginning with Session 18, they were informed that the sounds on the discrimination tapes were synthetic approximations to speech. At this point the labeling tapes were introduced, and for the next three sessions *Ss* were given four labeling tapes per session with instructions to identify each stimulus as a speech sound. During this period they were given no discrimination task. At Session 22 and thereafter *Ss* were asked to identify the stimuli on the labeling tapes as *b*, *d*, or *g*, and the discrimination task was resumed. In these sessions, the discrimination task was always undertaken after *Ss* had finished judging the stimuli on the labeling tapes.

For several reasons the discrimination data obtained before and after the introduction of the labeling tapes have been combined. First, an examination of the results showed that there was no obvious change in the discrimination judgments following the introduction of the labeling tapes and the instruction to judge the stimuli on those tapes as *b*, *d*, or *g*. Second, in the sessions in which the labeling tapes were first introduced, *Ss* reported that they had previously heard the sounds on the discrimination tapes as speech. Moreover, when they were asked to identify the speech sounds they heard, they responded mostly with *b*, *d*, and *g*.

When the discrimination data are combined, there are, for each *S* in Group I, a total of 21 judgments of each ABX triad. Since there were two triads for each combination of A and B stimuli, the total number of judgments of each A and B combination by each *S* was 42. For each of the stimuli on the labeling tapes there are, for each *S*, 32 judgments. All of the labeling judgments were obtained after Session 21—i.e., after *Ss* had been asked specifically to identify the stimuli on the labeling tapes as *b*, *d*, or *g*.

Two of the five *Ss* in Group I were eliminated because they failed to apply the phoneme labels consistently. Since these *Ss* did not clearly divide the stimulus continuum into phoneme categories, one cannot compare their discrimination of speech sounds within and across phoneme boundaries. It should be noted here that the stimuli of this experiment have previously been presented to large groups of listeners, and that it is quite unusual to find as many as two out of five who are as unreliable in their responses as the two *Ss* who were eliminated from this experiment.

Group II consisted of four workers at the Haskins Laboratories. These *Ss* had previously had extensive experience in listening to synthetic speech sounds. The procedure for Group II was

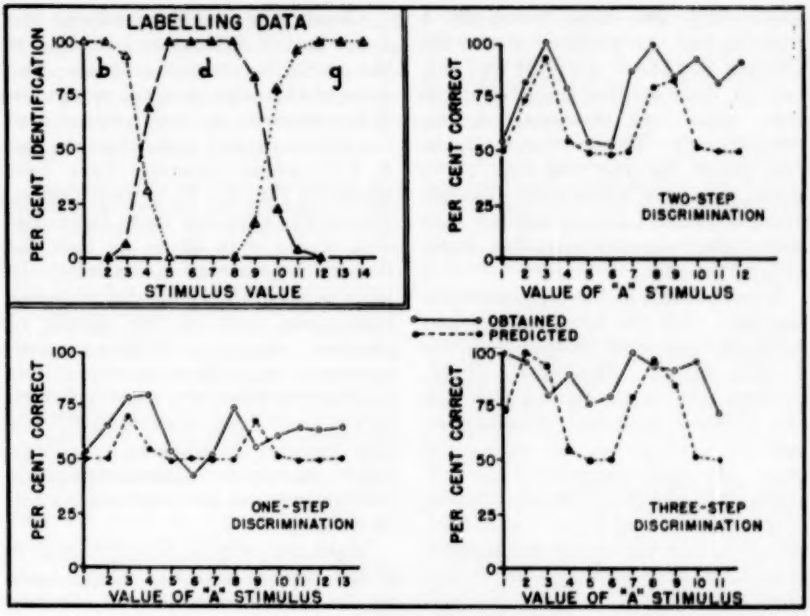

FIG. 2. Labeling and discrimination data for a single *S*, CD. The values given on the ordinates in terms of percentages are based on 32 and 42 judgments for the labeling and discrimination data, respectively.

like the procedure for Group I after Session 21. Thus, in each session, *S*s identified the stimuli on one labeling tape, and then judged three of the discrimination tapes—a total of 36 ABX triads. (This plan could not be followed exactly in the last few sessions due to scheduling difficulties.) In all, each *S* identified each of the stimuli on the labeling tapes 25 times, and he judged each ABX triad a total of 13 times. Since there were two triads for each A and B combination, each *S* judged each A-B combination a total of 26 times.

Of those *S*s whose data will be shown individually in the following section on results, CD and RV are from Group I, while LG and KH are from Group II.

RESULTS

In order to describe the method of presenting the data, and also to indicate in general terms the outcome of this study, the results that were obtained with one *S*, CD, will be presented first. At the upper left of Fig. 2 are plotted the labeling responses made by this *S* when the 14 stimuli were presented to him one at a time and in random order for judgment as *b*, *d*, or *g*. It can be seen that Stimuli 1, 2, and 3 were identified primarily as *b*, Stimuli 5, 6, 7, 8, and 9 as *d*, and Stimuli 11, 12, 13, and 14 as *g*. The shifts from one response to another are very abrupt, which is to say that the phoneme boundaries for this *S* are very sharp and stable.

The discrimination data obtained with the same *S*, CD, are also shown in Fig. 2. Only the "obtained" data points indicated by open circles and connected by solid lines will be considered at this time. Each point represents the percentage of correct responses for all ABX presentations

(both ABA and ABB) when the A stimulus had the value shown on the abscissa and the B stimulus was one, two, or three steps removed (for the one-, two-, and three-step curves, respectively). Thus, for example, the first point on the one-step curve shows that this *S* correctly discriminated Stimulus 1 and Stimulus 2 (one step higher on our stimulus scale) 54% of the time.

A comparison of the discrimination functions with the labeling functions indicates that, other things equal, this *S* does, indeed, discriminate better between stimuli that lie on either side of a phoneme boundary than he does between stimuli that fall within the same phoneme category. For example, it can be seen from the labeling curves that Stimuli 1 and 3 were both identified as *b* almost all of the time, and the two-step discrimination curves show that this *S* correctly discriminated these stimuli only 55% of the time. Two steps beyond Stimulus 3, which was almost always identified as *b*, is Stimulus 5, which was always identified as *d*. The discrimination curves show that these two stimuli, consistently labeled as different phonemes, were correctly discriminated 100% of the time.

It must, of course, be supposed that if a listener can always identify two stimuli as different (as in the case of Stimulus 3 and Stimulus 5), he will surely be able to discriminate them. On the other hand, it does not necessarily follow from the fact that he identifies two stimuli as the same phoneme (as with Stimulus 1 and Stimulus 3) that he cannot discriminate them. One might think that he would hear *two* types of *b*—i.e., that he would hear what the linguist calls allophonic variations. In the example cited, *S* does not.

Clearly, the data obtained with this *S* are not all so neat and striking as the particular examples chosen, and some of the other *S*s were more variable, especially in their responses to the discrimination task, than the one *S*, CD, whose responses have been shown in Fig. 2. It is, nevertheless, reasonably apparent from an inspection of the data of all *S*s that the discriminations tend to be relatively more acute in the vicinity of phoneme boundaries than in the middle of phoneme categories. Before presenting these other data, however, it is desirable to provide a basis on which all the results can be evaluated. For that purpose the working hypothesis will be stretched to a theoretical limit and its quantitative implications will be developed.

Make the extreme assumption that *S* can discriminate the stimuli only to the extent that he can identify them as different phonemes. Then suppose that in the discrimination task—i.e., when *S* is presented with the stimuli in ABX fashion and asked to say whether X is like A or like B—he can only assign the phonemic labels *b*, *d*, and *g* to the individual stimuli, and that he has no other basis for discriminating among the stimulus members of the various ABX triads. One can, then, use *S*'s responses in the phonemic labeling part of the experiment as a basis for calculating the frequency with which he will correctly discriminate in any given ABX arrangement of the stimuli. To do this, one must first refer to the labeling part of the experiment and take account of the relative frequency with which *S* identified each of the 14 stimuli as *b*, *d*, or *g*. It is, of course, possible to go from these data to calculations of the probabilities that a given ABX triad will be heard as any

one of the possible sequences of the three phonemes. One then needs only to determine for each triad which of the possible phonemic sequences will lead to responses that would be counted as correct discriminations.

By reference to the phonemic labeling part of the experiment we first determine for each A and B stimulus of the various ABX triads the relative frequency with which it was identified as *b*, *d*, and as *g*. These relative frequencies will be used as estimates of the probabilities of hearing the various stimuli of the ABX triads as *b*, as *d*, and as *g*. Let us call the probabilities of hearing these phones p_b, p_d, and p_g in the case of an A stimulus and p_b', p_d', and p_g' in the case of a B stimulus. We assume next that the various stimuli within each triad are perceived independently of each other. It follows, then, that the probability (for a given ABA triad) of hearing a particular sequence, such as *b*, *d*, *g*, is $p_b\,p_d'\,p_g$. Since there are three alternative responses, there are $3 \times 3 \times 3$, or 27, such phoneme sequences possible.

For any given triad the 27 possible phonemic sequences can be divided into three classes according to whether they lead *S* to make responses that would be counted as correct discriminations, as incorrect discriminations, or as discriminations that would on the average be correct half of the time and incorrect half of the time. In the case of any ABA type of triad, for example, *S* will be correct for any sequence in which the first and third stimulus members of the triad are heard as the same phoneme and the second member is heard as a different phoneme. He will be incorrect (again for the ABA type of triad) whenever the second and third members are heard as the same phoneme and the first is heard as a different phoneme. He will be correct half of the time and incorrect half of the time with two types of phoneme sequences: (*a*) all those in which he hears the first and second stimuli as the same phoneme and (*b*) those sequences in which he hears three different phonemes.[4]

[4] When *S* hears a sequence of three different phonemes, he can be expected to make the correct discrimination half of the time only if we assume that he perceives each phoneme as equally like the other two. It is possible that the typical listener does not perceive *b*, *d*, and *g* in precisely this way. However, the labeling data of this experiment are such as to produce essentially zero probabilities of hearing such sequences in the ABX triads. We have, therefore, not yet attempted to determine how the similarities among *b*, *d*, and *g* are perceived, because a correction, no matter how large, would have negligible effect on the results being reported here.

These considerations can be expressed quantitatively in the following way. Let $P\,Corr_{(ABA)}$ be the proportion of the time that the listener is correct on a number of presentations of the same ABA sequence, P_R be the proportion of the time that the listener heard a sequence which would lead to correct discrimination, P_W be the proportion of the time that the listener heard a sequence which would lead to incorrect discrimination, and P_I be the proportion of the time that the listener heard a sequence which would, with equal likelihood, lead to correct and incorrect discriminations. Then

$$P\,Corr_{(ABA)} = 1P_R + 0P_W + .5P_I.$$

If the probabilities of the particular sequences described above are substituted appropriately for the general expressions P_R, P_W, and P_I, and if the resulting equation is then manipulated algebraically, we obtain

$$P\,Corr_{(ABA)} = \frac{.5 + p_b^2 + p_d^2 + p_g^2 - p_b p_b' - p_d p_d' - p_g p_g'}{2}.$$

So far, we have been concerned only with the case in which the presented sequence is ABA; analogous considerations lead to a similar equation for ABB, although the particular sequences which are correct and incorrect are different.

$$P\,Corr_{(ABB)} = \frac{.5 + p_b'^2 + p_d'^2 + p_g'^2 - p_b p_b' - p_d p_d' - p_g p_g'}{2}.$$

The ABA and ABB sequences were presented equally often; therefore, an average $P\,Corr$ is given by

$$P\,Corr = \frac{.5 + (p_b - p_b')^2 + (p_d - p_d')^2 + (p_g - p_g')^2}{4}.$$

The solid points connected by dashed lines in Fig. 2 represent the discrimination values derived from CD's labeling curves on the assumptions outlined above. It is seen in the case of this *S* that the predicted functions do indeed take care of some, although by no means all, of the

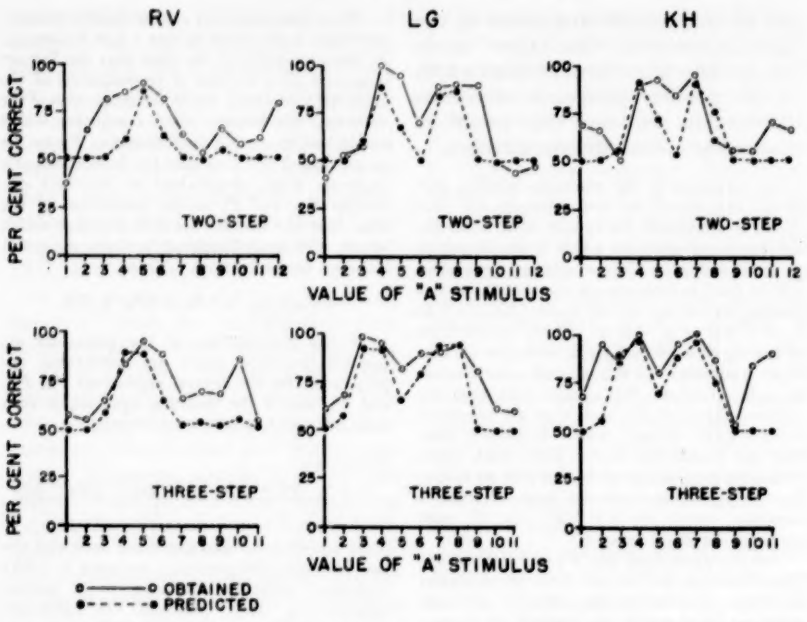

FIG. 3. Predicted and obtained discrimination values at two and three steps for RV, LG, and KH. Each of the obtained values is based on 42 judgments in the case of RV, and on 26 judgments for LG and KH.

variations in the obtained discriminations. The assumptions predict the points of high and low discrimination reasonably well, but they lead one to expect a general level of discrimination slightly lower than that obtained.

In Fig. 3 are data obtained with three other *S*s. These data are not presented as a way of indicating the results for all *S*s but only to show additional details of the results, including in particular a sample of the individual differences among *S*s. It is seen that the position and number of the peaks in the predicted discrimination functions vary somewhat from one *S* to another, reflecting differences in the way *S*s had assigned phoneme labels to the stimuli when they were presented for identification. It is also apparent that the obtained discriminations follow the inter-*S* differences in the predicted functions fairly well.

The simplest way to summarize the data for all *S*s is to make a scatter plot of obtained values against predicted values. Such plots are shown in Fig. 4, 5, and 6, for one-, two-, and three-step discriminations, respectively.[5] For the two- and three-step data, regression lines have been fitted by the method of least squares, and these are shown for each set of points. The regression for the one-step data has not been determined because so

[5] It should be noted here that the reliabilities of all points in the scatter plots are not equal, since, as was pointed out under the section on procedure, the two groups of *S*s made different numbers of judgments of the various ABX triads.

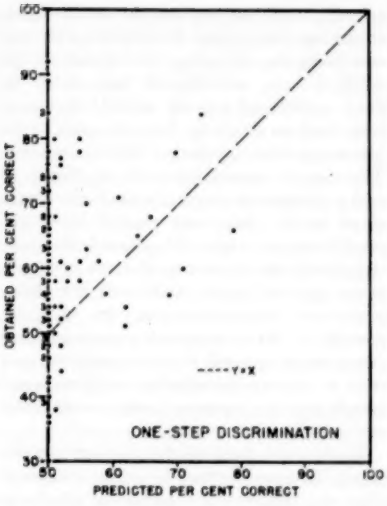

FIG. 4. Scatter plot of predicted vs. obtained values on the one-step discrimination for all *Ss*. The small numerals have been placed on the graph to indicate, where necessary, the number of values that occupy the same position of the coordinates.

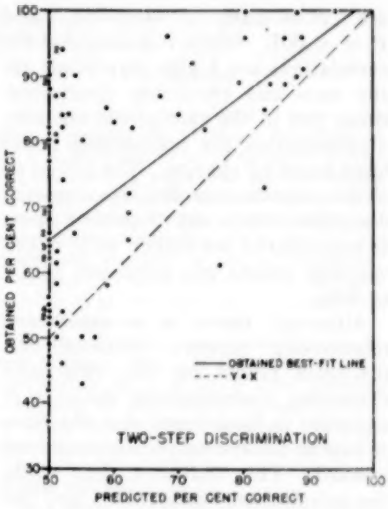

FIG. 5. Scatter plot of predicted vs. obtained values on the two-step discrimination for all *Ss*. The small numerals have the same meaning as in Fig. 4.

few predicted points lie above 50% that a meaningful fit cannot be obtained. The one-step data do not, therefore, provide a good test of our assumptions. For the two- and three-step data we should, of course, suppose that if the obtained data were essentially as predicted, give or take a little experimental error, the regression lines would be described by the equation "$x = y$."

The relationship between predicted and obtained values has been measured by computing tau,[6] a nonparametric measure of correlation. The correlations are +.14, +.43, and +.43 for the one-, two-, and three-step discriminations, respectively. Significance levels, in the same order,

[6] We have used a nonparametric measure because our data fail to meet the assumption of homoscedasticity. For a description of tau, see Kendall (5).

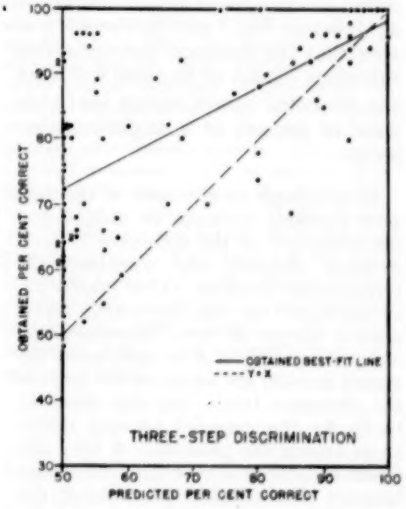

FIG. 6. Scatter plot of predicted vs. obtained values on the three-step discrimination for all *Ss*. The small numerals have the same meaning as in Fig. 4.

are $P = <.08$, $P = <.001$, and $P = <.001$. Thus, it is seen that the correlations are highly significant for the two- and three-step discriminations, but in the case of the one-step discrimination the relationship could have arisen by chance. The failure to obtain significance for the one-step discrimination is not surprising since, as was pointed out earlier, most of the one-step points are predicted to lie at 50%.

Although there is a significant relationship between obtained and predicted points for the two- and three-step discrimination data, it is apparent in both cases that the lines of best fit are systematically displaced upward. This indicates that while the assumptions predict fairly well the occurrence and location of the inflections in the discrimination curves, they apparently lead to an underestimation of the general level of discrimination. This was previously noted in the data of the individual *S*s as shown in Fig. 2 and 3, where it was seen that the obtained discrimination functions tended to fit quite well with the predicted curves except that they were in general at a slightly higher level.

It is difficult on the basis of the data now available to make an unequivocal interpretation of the difference in level between obtained and predicted discrimination functions. One possibility, of course, is that this discrepancy represents a margin of "true" discrimination —i.e., an ability of *S* to distinguish the speech sounds, not solely on the basis of the phonemic labels, but also more directly by the essential acoustic differences among the patterns. A very different possibility is that the discrepancy between obtained and predicted is the result of certain detailed aspects of the experimental procedure. For example, irrelevant discriminable aspects of the stimuli, such as accidental stray noise, could have provided *S*s with an extraneous basis for deciding, in regard to the ABX triads, whether X was A or B. Such additional stimuli would, of course, have had no effect on *S*s' responses in the phonemic labeling part of the experiment. The result would have been, then, to make the obtained discriminations somewhat better than one should have expected them to be. In general, the procedures of this experiment were not such as to control most effectively for these irrelevant discriminanda. It will be possible in future research to take greater precautions against their occurrence, and then to determine whether and to what extent the discrepancy between obtained and predicted is reduced.

Within the data of the present experiment, however, there is some evidence that the discrepancy between obtained and predicted discrimination functions is due, at least in part, to "true" discrimination. This evidence is to be found in the fact that the discrepancy would appear to be greater for the three-step than for the two-step function. One might expect to obtain this difference if *S* were truly discriminating the stimuli on the basis of their essential acoustic characteristics, since one would then presumably find as between the two- and three-step discriminations that the three-step discrimination was the easier. Other factors, such as the irrelevant discriminanda discussed in the preceding paragraph, would have affected the two- and three-step conditions equally, and would not have caused the departure from predicted values to be greater in the three- than in the two-step data.

Discussion

The results of this experiment cannot be assumed unequivocally to reflect the effects of learning on discrimination. There is, of course, the possibility that the inflections in the discrimination function are given innately, and that the phoneme boundaries have been placed

so as to coincide with these discontinuities. This begins to seem unlikely when one considers that other languages have put phoneme boundaries at different places along the *b-d-g* continuum. In itself, this would appear to reduce the probability that the sharp inflections in the discrimination functions are innately given. One would have far more compelling evidence, however, if it were found that native speakers of such languages have their points of maximal and minimal differential sensitivity displaced along the continuum to correspond with the phoneme boundaries of their respective languages. Research to test this possibility is feasible.

In order to find out whether the effects here described represent acquired similarity, acquired distinctiveness, or a combination of both, it will, of course, be appropriate to obtain discrimination data on nonspeech stimuli that are otherwise identical with the synthetic approximations to speech. Complete identity is obviously impossible, unless one can get *Ss* to hear the same sounds sometimes as speech and sometimes not. One can, however, reasonably expect to make revealing comparisons between discrimination data obtained with speech and nonspeech stimuli that vary along the same, simple stimulus dimensions. Sheer temporal duration, for example, is sometimes a cue for distinguishing speech sounds. (Duration of transitions distinguishes stop consonant from semivowel; duration of fricative noise is a cue for distinguishing among the classes fricative, affricate, and stop.) It will be possible to obtain discrimination functions for variations in duration when those variations cue the perceived differences among phonemes, and to obtain comparable data for variations in duration of nonspeech sounds. We should suppose that a comparison of discrimination functions such as these would help greatly to determine whether the typical listener's long training in speech perception has served selectively to sharpen or to dull his discrimination of speech sounds, or whether, perhaps, it has done both.

Summary

This experiment was designed to measure the relation between *Ss'* phonemic identifications (as *b*, *d*, *g*) of certain synthetic speech sounds and the extent to which they can discriminate the sounds as being different in any way. The stimuli were two-formant approximations to consonant-vowel syllables. They varied in the extent and direction of the second-formant transitions, this variable having previously been found to be an important cue for the perceived distinctions among *b*, *d*, and *g*.

In one part of the experiment these sounds were presented singly and in random order for identification as *b*, *d*, or *g*. The responses obtained in this way tended, with most listeners, to divide the stimulus continuum into three sharply bounded phonemic categories, indicating that the perceived shifts from one phoneme to another were rather abrupt.

In a second part of the experiment an ABX procedure was used to measure *Ss'* ability to discriminate these sounds. The results indicated that discrimination was better at the phoneme boundaries than in the middle of the phoneme category. That is, with acoustic differences equal, *Ss* discriminated better between speech sounds to which they habitually attach different phonemic labels than they did between sounds which they normally put into the same phoneme class. To obtain a more nearly precise evaluation of this effect, the obtained discrimination curves were compared with those that would be predicted from the phoneme identification curves on a most extreme assumption: namely, that *S* can only hear those differences that are revealed by the way in which he attached phonemic labels to the stimuli. The discrimination curves that were produced on this assumption predicted fairly well the occurrence and location of points of high and low discriminability in the obtained functions, but *Ss* tended in general to discriminate the stimuli somewhat better than expected from the extreme assumption.

References

1. Cooper, F. S. Spectrum analysis. *J. acoust. Soc. Amer.*, 1950, **22**, 761–762.
2. Cooper, F. S. Some instrumental aids to research on speech. In *Report of the fourth annual round table meeting on*

linguistics and language teaching. Washington, D. C.: Institute of Languages and Linguistics, Georgetown Univer., 1953. Pp. 46–53.

3. Cooper, F. S., Liberman, A. M., & Borst, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Nat. Acad. Sci.*, 1951, **37**, 318–325.
4. Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, 1955, **27**, 769–773.
5. Kendall, M. G. *Rank correlation methods*. London: C. Griffin, 1948.
6. Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.*, 1954, **68**, No. 8 (Whole No. 379).

(Received November 6, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

RETENTION OF VERBAL ASSOCIATES AS A FUNCTION OF STRENGTH[1]

WILLARD N. RUNQUIST[2]

Northwestern University

This study was concerned with some basic relationships between interfering responses and forgetting of items in a paired-associate verbal list. In the retroactive and proactive inhibition designs, where the sources of interference can be at least partially specified, retention has been assumed to be a function of the relative strengths of the competing responses at the time of recall. Supporting this assumption are the studies varying degrees of first-list learning (**6, 14**), degree of second-list learning (**1, 7, 8, 9, 11, 12**), or both (**3**), which have shown an increase in retroactive inhibition (RI) and proactive inhibition (PI) as the degree of learning of the interfering list increases and a decrease in RI as the degree of learning of the recalled list increases. In this study an attempt was made to analyze the retention of the individual items within the list as related to their strength.

Two relationships were studied. The first of these was the relation between retention and the strength of the competing responses when the two responses are learned to the same stimulus. Recall of both first- and second-learned responses was studied since some evidence has indicated differences in recall due to the temporal position of the interfering learning (**13**). The second relation investigated was that between forgetting and item strength when there was no specific interfering learning. It is known at a gross level that response strength influences retention, although the exact nature of the relationship is not known.

Method

Procedure.—Three groups of 104 *Ss* each learned lists of eight paired adjectives. Only *Ss* who had not served in verbal learning experiments were used in order to reduce proactive interference from outside this particular experiment to a minimum.

Four sets of two lists were taken from a previous study by Underwood (**13**). These lists originally contained 10 pairs so two pairs were dropped arbitrarily from each list. The stimuli of the two lists within a set were identical, with the responses being different. Intralist and interlist similarity were low. Each *S* learned lists from a single set with an equal number of *Ss* in each group learning each list in each set first. The purpose of using several lists was to avoid biasing the results with specific item effects as much as possible. The three groups were designated RI, PI, and Control. The control group learned and recalled a single list, while the RI and PI groups learned the two lists of a set 15 sec. apart, with the RI group recalling the first list and the PI group the second list. A recall trial and four additional relearning trials were given after 24 hr. An attempt was made by careful instruction to insure that *Ss* knew exactly which list they were to recall in the RI and PI conditions.

Lists were presented on a Patterson S-PA Memory Drum at a 2:2-sec. rate with a 4-sec. intertrial interval. All learning was by the anticipation method to one perfect trial plus two additional trials. Five orders of presentation of the pairs were used to minimize serial learning.

Measurement of response strength.—There are several ways in which the response strength of the items could be dimensionalized. The most obvious way is in terms of the number of correct anticipations (reinforcements) on each item. If the eight items for each *S* are rank ordered according to the number of reinforcements, these ranks can be used as an ordinal dimen-

[1] This study is abstracted from a doctor's dissertation submitted to the Graduate School, Northwestern University. The author is indebted to Dr. Benton J. Underwood for his supervision.

[2] Now at the State University of Iowa.

sion of response strength and retention of items for each rank can be plotted. This method of dimensionalizing strength has the advantage of having each *S* contributing one item to each point on the dimension. It was selected over several other possible methods because the alternative methods appeared to give biased estimates of the amount of forgetting.

If the raw percentage of items recalled for each rank were used as a measure of forgetting, this would not give a true picture of this relationship. If, for example, recall were given immediately by having an additional learning trial, the items of Rank 1 (large number of reinforcements) would be correctly recalled more often than items of Rank 8. Thus, a measure of forgetting must only take into account the loss of items beyond that obtained with immediate recall if it is not to reflect merely the learning differences. This problem has been more fully discussed elsewhere (**15**). An estimate of the expected percentage recalled on the trial immediately following the completion of learning may be obtained by successive probability analysis (**15**). Briefly, this method consists of taking each item for each *S* and tracing it through the course of learning noting whether it was correct or incorrect following each reinforcement. By pooling the data over items and *S*s, the percentage of items correct on the next trial following 1, 2, 3, etc. reinforcements is obtained. These percentages can be used as estimates of the probability that a certain item will be correctly anticipated on an immediate recall trial by noting the total number of reinforcements and looking up the value of the percentage.

The more homogeneous the items which are pooled, the more accurate will be this estimated probability. Therefore, separate probability analyses were performed as described above for fast and slow learning *S*s as determined by dividing the distribution of trials to criterion at the median. Separate analyses were also performed within each category for items of Ranks 1 and 2 combined, Ranks 3 and 4 combined, Ranks 5 and 6 combined, and Ranks 7 and 8 combined. Thus the predicted probability (percentage) of an item being correctly anticipated at immediate recall was determined only from items of the same approximate strength from *S*s with approximately the same speed of learning.

Retention was measured in terms of absolute loss in probability or percentage. The obtained percentage recalled for each rank was subtracted from the predicted percentage. The predicted percentage was obtained by averaging the predicted probabilities for each item obtained from the analysis described above. Whether or not this gives an accurate picture of forgetting depends upon certain assumptions concerning the relationship of forgetting to strength. By using the absolute loss, the assumption is made that this relation is linear. Thus a 50% loss would be treated as an equal amount of forgetting whether the predicted recall was 90% or 60%. Another and possibly intuitively more plausible method, would be to measure forgetting in terms of relative amount forgotten. This is done by taking the percentage of the expected recall which is obtained. In these data, the results obtained by both methods were almost identical, hence the absolute measure will be used as it is more convenient to handle statistically.

Results

Retention of items in a single list.— In determining a relationship between response strength and forgetting, only the data from the control group are considered. The mean number of items recalled per *S* after 24 hr. was $5.83 \pm .12$ which represents 73% recall. The absolute loss in probability for the eight ranks is shown in Fig. 1 with Rank 8 representing the weak items and Rank 1 the strong. The mean predicted probabilities for the eight ranks were .62, .75, .85, .87, .92, .96, .97, and .99 from Rank 8 to Rank 1, respectively. While it may not be legitimate to space the ranks equally, the nature of the relationship which shows that forgetting first in-

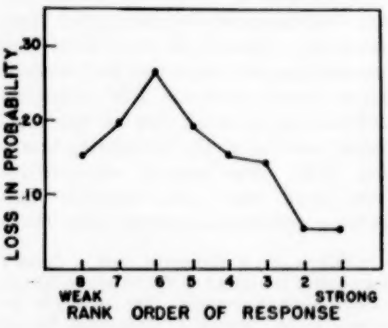

Fig. 1. Absolute loss in probability as a function of the rank order of the response.

creases then decreases is not affected. In fact, the increase in forgetting for the weaker items is more extended if the points are plotted with the predicted probabilities on the abscissa.

Statistical evaluation of the relationship required obtaining a score for each item for each *S*. This involved subtracting the obtained percentage recalled from the predicted percentage recalled for each item. Since a single item is either recalled (100%) or not recalled (0%), 100 was subtracted from the predicted percentage (probability) of each item correctly recalled, while zero was subtracted from the predicted percentage of each item not correctly recalled. Actually, these loss scores are artificial in terms of percentage. What was desired was to assign a value to each item in such a way that *S* would receive more credit for getting a weak item correct than he would for getting a strong item correct. The loss scores are thus weighted recall scores for each item. The mean of these loss scores for each rank corresponds exactly with the over-all loss as determined above. Since the distributions of loss scores were bimodal, a Friedman chi-square (**10**) was used to test for differences between the ranks. The resultant chi square was 74.07 which was highly significant ($P < .001$ for 7 df). Further evidence for the reliability of this relationship is that the same general function has been obtained by Underwood in two other independent instances, although these data have not been published.[3]

In view of this unusual relationship one might ask whether some procedural artifact might have produced the results. If there is such an artifact it could not be discovered. Several methods of categorizing items as to strength produced similar shaped functions, once the known sources of bias were taken into account. These biases were usually in the form of either distorted estimates of the predicted probabilities or selection of items for particular points from certain kinds of *S*s such as fast and slow learners. Therefore, within the limits of this study, this appears to be the relationship between the strength of the association and the amount of forgetting. It should be pointed out that there is no a priori reason why the interaction of interference within the list, which is presumably responsible for the differential forgetting of various items, should produce the same law as when response strength is varied by having lists learned to a different criterion. A systematic study of the influence of variables affecting interference, such as similarity, would appear to be necessary before a satisfactory account of the phenomenon reported here could be made.

RI and PI.—The mean number of items recalled under RI and PI conditions were 3.36 ± .13 and 3.41 ± .17, respectively. All items which were in the recalled list for these two groups were rank ordered in terms of the number of reinforcements on that item, and independently on the number of reinforcements on the corresponding interfering item. A two-way classification was then made of all items according to the rank order of the recalled item (List 1 for the RI group and List 2 for the PI group) and the rank order of its corresponding interfering item. Ranks 1 and 2, 3 and 4, 5 and 6, and 7 and 8 were combined, in order to achieve better reliability by including more items at each point. This gave a 4 × 4 classification for each group. In order to measure the amount of RI and PI at each point on the strength dimension, it was first necessary to determine the loss (forgetting) in probability or percentage at each point corrected for differences in initial strength exactly

[3] B. J. Underwood, personal communication, October, 1956.

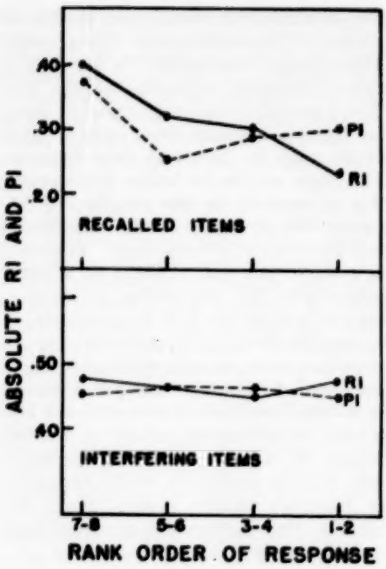

FIG. 2. Absolute RI and PI as a function of the rank order of the recalled items and the rank order of the interfering item.

as in the control group data. The amount of absolute RI and PI was then determined by subtracting the loss which took place without specific interfering learning; i.e., the loss in probability for the control group at that point. RI and PI calculated in this way are plotted in Fig. 2. The upper function gives RI and PI as a function of the rank of the recalled item ignoring the rank of the interfering items, while the lower function gives RI and PI as a function of the rank of the interfering item ignoring the rank of the recalled item.

For purposes of statistical evaluation a score was obtained for each *S* at each point by subtracting the obtained percentage recalled from the predicted percentage recalled, and then subtracting the control group's percentage recalled for the same point. Since each *S* contributed two items to each point, the predicted percentage recalled was the average of that for these two items. The obtained percentage recalled was either 100, 50, or zero depending upon whether *S* recalled both, one, or neither of the items. Again the mean RI and PI computed in this way corresponded to the curves obtained by the over-all analysis. The RI and PI scores were analyzed by means of four Friedman chi-square tests. The first of these tested RI as a function of the rank of the recalled items. The chi square was 21.05 which was significant beyond the .001 level with 3 *df*. The second tested PI as a function of the rank of the recalled item. The chi square was 5.26 which was not significant ($P = .14$). The other two tested RI and PI respectively as a function of the rank of the interfering item. These values were .84 and 2.40, neither of which approached significance.

As expected, the stronger the item, the less the RI. This result corresponds quite well with studies in which the degree of learning was varied by manipulating the number of trials. The strength of the item did not appear to be an effective variable in the amount of PI, although there is some suggestion of a decrease in PI as strength increases. The difference between RI and PI may possibly be attributed to a difference in the range of total reinforcements. Since there was over all some positive transfer in learning List 2, the range of reinforcements was less on List 2 than List 1. Although there was no difference in the predicted probabilities for the corresponding ranks between the RI and PI groups, there is evidence which indicates that there is an increase in retention even when probabilities do not increase any further with the number of reinforcements.

It is possible that the restriction in range of number of reinforcements in the PI group was responsible for no relationship being obtained. It is impossible to compare this result with the amount of PI produced by varying numbers of trials on the second list, since no studies have been done using this variable. Data by Briggs (**3**) shows that *Ss* given free recall after learning two lists to various criteria do not show large changes in the amount of PI with some degrees of first-list learning.

The fact that strength of the interfering item is not related to RI and PI is somewhat surprising in view of the vast number of studies which show that the lower the degree of learning of the interfering list, the less the RI and PI. The result obtained here suggests that the differentiation between the two response systems may be the most important factor in forgetting due to interfering learning. Support for this notion also comes from recent studies by Bilodeau and Schlosberg (**2**) and Greenspoon and Ranyard (**5**) which show considerably less interference when the two lists are learned in distinctive environments.

These functions do not appear to be the result of any interaction between items of different strengths. When plotted as a joint function of the rank of the recalled and interfering item, RI and PI did not show any systematic variation or interaction other than the relationships previously discussed. These data will not be presented.

The important fact obtained from the retention and RI and PI data is that the laws obtained between strength of association and retention when strength is measured within a list are not the same as those obtained when strength is varied by giving *Ss* different numbers of trials on the lists as a whole.

Relearning.—Further evidence for a difference in RI and PI occurs in the relearning data. The mean number of correct responses on each of the five relearning trials is plotted in Fig. 3. Analysis of variance on the mean total correct responses in relearning gave an F of 65.66 with heterogeneous variances due to the restriction of the range in the control group. A t test between the RI and PI groups gave a value of 2.18 ($P = .03$). The difference occurs largely on the last two trials where the PI group performs as well as the control group, while the RI group is somewhat retarded. It has been the usual finding that significant RI and PI disappear after two relearning trials (**9, 11, 12, 13, 14**). The present study gave significant RI throughout the five relearning trials and significant PI up to the fourth trial. The studies cited above are not strictly comparable since the *Ss* served in all conditions in those studies.

The overt error data suggest that this difference in RI and PI may be attributed to intralist rather than interlist interference. From Table 1, it can be seen that the RI and PI groups made more total errors during relearning than the control group ($t = 5.19$ and 3.05 for the RI vs. control and PI vs. control, respec-

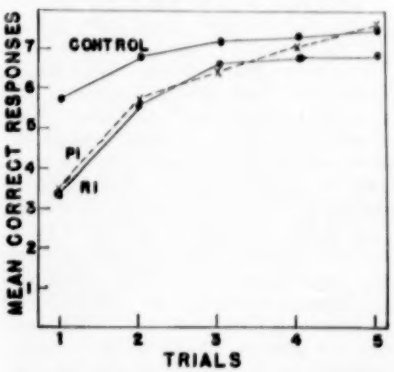

FIG. 3. Mean number of responses correct on five relearning trials.

TABLE 1

MEAN ERRORS PER *S* FOR THE FIVE RELEARNING TRIALS

Group	Total Errors		Intrusions	
	Mean	σ_M	Mean	σ_M
Control	2.14	.22	—	—
RI	3.94	.27	1.05	.16
PI	3.13	.24	1.26	.13

tively). The RI group also made more total errors than the PI group ($t = 2.26$, $P = .03$), but this difference seems to be due to a greater number of intralist errors since the number of intrusions did not differ between the two groups. A further analysis of the errors occurring in relearning showed that the RI group made a greater number of repeated errors and errors previously made in original learning, while most of the PI errors were intrusions or had never occurred before. In terms of Gibson's hypothesis (**4**) concerning the spontaneous recovery of generalization it appears as if the RI group had a more difficult time eliminating intralist generalization tendencies recovered from the original learning.

The effects of item difficulty.—In this study, the strength of an item and its difficulty are, of course, confounded. The above relationships, then, should best be described in terms of a joint function of strength and difficulty. This is not crucial as it is precisely what the study set out to discover. It was originally intended to also analyze transfer to the learning of the List 2 items as a function of the strength of the List 1 item. Several analyses were attempted, with the general conclusion being that the relationship could not be obtained unambiguously from these data. This was due to a possible correlation between the difficulty of the List 1 item and the difficulty of the List 2 item. This could be perhaps due to the greater distinctiveness of some stimulus components. Thus any measure of the learning of a List 2 item would be confounded with the difficulty of that item. This would not, however, enter into the recall data, since the interest here is not in the learning of the List 2 item, but only in its empirical strength at the end of learning regardless of how or why it is of that strength.

SUMMARY

The purpose of this study was to determine the relationship between retention and the strength of the association, and the relationship between RI and PI and the strength of the competing responses.

Two groups of 104 *Ss* each learned two paired adjective lists with identical stimuli and different responses. A third group of 104 *Ss* learned a single list. After 24 hr., all *Ss* recalled and relearned either the first list (RI) or second list (PI).

An item analysis was performed on the data. Items were ranked according to the number of reinforcements and retention scores were obtained for each rank corrected for the strength of the response at the end of learning. The results showed:

1. Forgetting of the items in a single list increased then decreased as the strength of the item increased.
2. RI was a function of the strength of the recalled item, but was not related to interfering item strength. PI was not related to either recalled item strength or interfering item strength.
3. Relearning was faster for the PI group than for the RI group.

REFERENCES

1. ARCHER, E. J., & UNDERWOOD, B. J. Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *J. exp. Psychol.*, 1951, **42**, 283–290.
2. BILODEAU, I. M., & SCHLOSBERG, H. Similarity in stimulus conditions and retroactive inhibition. *J. exp. Psychol.*, 1951, **41**, 199–204.
3. BRIGGS, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.
4. GIBSON, E. J. A systematic application of the concepts of generalization and differ-

entiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.

5. Greenspoon, J., & Ranyard, R. Stimulus conditions and retroactive inhibition. *J. exp. Psychol.*, 1957, **53,** 55–59.
6. McGeoch, J. A. The influence of degree of learning on retroactive inhibition. *Amer. J. Psychol.*, 1929, **41,** 252–262.
7. McGeoch, J. A. The influence of degree of interpolated learning on retroactive inhibition. *Amer. J. Psychol.*, 1932, **44,** 695–708.
8. McGeoch, J. A. Studies in retroactive inhibition: VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated list. *J. exp. Psychol.*, 1936, **19,** 674–693.
9. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.
10. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.
11. Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32,** 185–200.
12. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59,** No. 3 (Whole No. 273).
13. Underwood, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38,** 29–38.
14. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39,** 24–34.
15. Underwood, B. J. Speed of learning and amount retained: a study in methodology. *Psychol. Bull.*, 1954, **51,** 276–282.

(Received November 16, 1956)

Journal of Experimental Psychology
Vol. 54, No. 5, 1957

SUPPLEMENTARY REPORT: EFFECT OF VERBAL PRETRAINING ON THE ACQUISITION OF A COMPLEX MOTOR SKILL[1]

WILLIAM F. BATTIG,[2] DONALD R. HOFFELD, SIDNEY SEIDENSTEIN, AND W. J. BROGDEN

University of Wisconsin

Battig (**3**) found that amount of positive transfer from verbal pretraining to performance of a finger-positioning task decreased as complexity increased from one to two fingers and was zero for the 3- and 4-finger tasks. These and other results (**2**) suggest that relevant verbal pretraining has little or no effect on the acquisition of complex motor tasks. Insufficient relevance of the verbal pretraining used in previous studies could account for these negative results. The present study is an attempt to increase the degree of relevance of verbal pretraining in order to test that hypothesis, with Battig's (**3**) study as the point of departure.

Method.—Battig's (**3**) finger positioning apparatus was used with the 4-finger task. The *S*'s were 36 University of Wisconsin psychology students assigned at random to three groups of 12 each. Each of three *E*'s ran four *S*'s in each group. The three groups differed only in type of pretraining. Group I, the control group, had no pretraining. For Group II—the stimulus-words pretraining group—the procedure and letters were identical to those for Battig's (**3**) relevant-*S* group, the difference being that the letters could be seen only when the lights were lit. For Group III—the response-words pretraining group—a letter of the alphabet was placed over each response light, so that it was seen only when the light was lit by appropriate positioning of the response key. The letters were the same as for Group II, and *S*'s task was the same: to pronounce nonsense words formed from the letters covering the green lights that corresponded to the red lights of the stimulus pattern. Group III differed from Group II primarily in that *S* could identify the letters only

[1] Supported in part by a grant from the National Science Foundation.

[2] Now at Psychological Laboratory, University of Virginia.

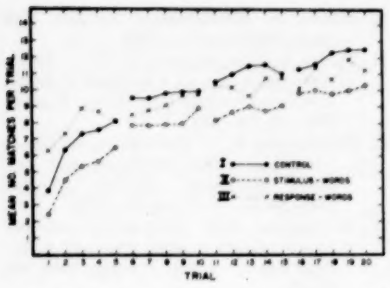

FIG. 1. Mean number of matches per trial for each group. Breaks in curves represent 1-min. intertrial rest.

by moving the response keys. However, *S* was not instructed to match the red and green lights, only to move the keys until the correct letters could be identified.

The instructions given each *S* and the conditions and procedure during the experimental session were those used by Battig (**3**).

Results.—The mean number of matches per trial for the three groups is presented in Fig. 1. Analysis of variance of scores summed over 20 trials shows trials and the interaction of trials by groups to be the only significant sources of variation. For analysis of variance of the data for the first trial and the first five trials, a significant difference between groups occurs in both cases. The Duncan range test (**4**) shows that on Trial 1, Group III is significantly superior to both Groups I and II, which do not differ from each other. For Trials 1–5, Group III is significantly superior to Group II, but neither differs from Group I. Analysis of variance of Trials 6–20 yields no significant differences between groups.

An Alexander trend test (**1**) on the group curves provides further analysis of the Trials × Groups interaction. Significant over-all slope and over-all deviations from linearity show significant nonlinear acquisition. The over-all between-groups slope term is significant, but pairwise comparisons show only that the slope for Group III is significantly less steep than the slopes for Groups I and II.

Discussion.—The pretraining for Group III involved a form of motor practice (manipulating the response keys in order to view the letters on the stimulus lights) that may account for the initial positive transfer. The amount of practice in pretraining for Group III is roughly equivalent to the amount of practice for Group I on Trial 8. A t test of the difference between Group I on Trial 8 and Group III on Trial 1 shows Group I to have a statistically significant superior performance. The difference between Groups II and III is probably due also to the added motor practice given Group III during pretraining and provides no evidence of a difference in the effect of the verbal factors involved in the two procedures. Thus, the results of this experiment give little or no evidence for positive transfer from either type of verbal pretraining and support Battig's (**3**) conclusion that verbal pretraining has no beneficial effect on acquisition of a motor task as complex as the 4-finger task used here.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. BATTIG, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *J. exp. Psychol.*, 1954, **47**, 371–380.
3. BATTIG, W. F. Transfer from verbal pretraining to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51**, 371–378.
4. DUNCAN, R. B. A significance test for differences between rank treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2** (n. s.), 171–189.

(Received October 11, 1956)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band IV/2

CARL-FRIEDRICH GRAUMANN, **Zur Phänomenologie des Testverfahrens.**

REINHARD TAUSCH, **Nicht-directive Erwachsenen-Psychotherapie bei deutschen Klienten.**

HELMUT SEYFRIED, **Unauffällige und schwererziehbare Knaben im Farbpyramidentest (nach Max Pfister).**

JOHANNES VOIGT, **Über individuelle Varietäten der Ermüdung.**

KURT MÜLLER, **Über visuell wahrgenommene Geschwindigkeit sagittal gerichteter Bewegungen.**

GUSTAV A. LIENERT, **Ein statistisches Konzept zur präzisen und rationellen Interpretation von Testprofilen.**

KLAUS-DIETER HARTMANN, **Kritische Buchbesprechung: J. Linschoten, Strukturanalyse der binokularen Tiefenwahrnehmung.**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.— per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
(Brentanoweg 10, Germany)

ARE THERE GAPS IN YOUR FILES OF APA JOURNALS?

Then hear this . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all available issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

American Psychologist

Journal of Abnormal & Social Psychology

Journal of Applied Psychology

Journal of Comparative & Physiological Psychology
(1947–1950 only)

Journal of Consulting Psychology

Journal of Experimental Psychology

Psychological Abstracts

Psychological Bulletin

Psychological Index
(a few complete volumes, some shopworn)

Psychological Monographs

Psychological Review

Not all issues in all volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery:
6 to 8 weeks

No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

Order from:

American Psychological Association
Department BB
1333 Sixteenth Street, N.W.
Washington 6, D. C.

SALE & BACK ISSUES